



Universität Regensburg

# **Situationsabhängige Rezeption von Information bei Verwendung der Wikipedia**

Inaugural-Dissertation zur Erlangung der Doktorwürde der Philosophischen Fakultät III  
(Sprach- und Literaturwissenschaften) der Universität Regensburg

vorgelegt von:

Hanna Knäusl

aus

93053 Regensburg

Regensburg, 2014

Erstgutachter: Prof. Dr. Rainer Hammwöhner  
Zweitgutachter: Prof. Dr. Christian Wolff

## ***Danksagung***

Die vorliegende Arbeit wurde im Wintersemester 2014/2015 als Dissertation angenommen. Mein Dank gilt ALLEN, die mich bei der Erstellung dieser Arbeit sowohl in der Form von fachlichem Austausch als auch allgemein motivierend unterstützt haben.

Folgenden Personen möchte ich besonders danken:

Meinem Doktorvater, Prof. Dr. Rainer Hammwöhner, der mich nicht nur mit fachlicher sondern auch menschlicher Unterstützung bei der Erstellung dieser Arbeit begleitet hat. Es gelang ihm, bei der Betreuung eine gute Balance zu finden. So konnte ich meine eigenen Ideen in der Forschung unterbringen, hatte aber gleichzeitig eine sehr kompetente Begleitung für die Erstellung einer stringenten wissenschaftlichen Arbeit. Bei Prof. Dr. Christian Wolff bedanke ich mich für seine Bereitschaft, als Gutachter für meine Arbeit zu fungieren.

Ich möchte mich auch bei meinen Kollegen bedanken, die durch ihre Bereitschaft zu fruchtbaren Diskussionen und zu konstruktivem Feedback ebenfalls zum Gelingen dieser Arbeit beigetragen haben.

Für die stets motivierende Anteilnahme möchte ich mich bei allen meinen Freundinnen, vor allem Lea, Diana, Steffi und Iris, bedanken.

Ein überaus großer Dank gebührt auch meiner Familie: Meinen Eltern allgemein für ihre immer vorhandene Unterstützung und ganz speziell für das Korrekturlesen dieser Arbeit. Meinen Söhnen Johannes und Julius möchte ich danken, dass sie mir durch ihre unkomplizierte Art die Durchführung des Dissertationsprojekts ermöglicht haben. Und nicht zuletzt danke ich ihrem Vater Sebastian, der mit sehr viel Geduld und Sachlichkeit gerade in schwierigen Phasen eine große Stütze war.

Regensburg im Dezember 2014

*Hanna Knäusl*

## ***Zusammenfassung***

Diese Arbeit beschäftigt sich mit der Frage nach einer situationsrelativen Informationsrezeption im Information Retrieval im Allgemeinen und bei der Suche in der deutschsprachigen Wikipedia im Speziellen. Motiviert wurde das Forschungsdesiderat vor allem durch die schnell ansteigenden Datenmengen und die damit verbundenen Selektionsschwierigkeiten durch den Nutzer sowie weiter durch die Informationssuche als Beschäftigung des alltäglichen Lebens.

Die Arbeit ist dabei in zwei Teile untergliedert. Im theoretischen Teil werden zunächst die Komplexität der möglichen Suchsituationen und der Einfluss subjektiver Komponenten auf die Bewertung von Informationen besprochen. Auf dieser Basis wurde anschließend das situationsabhängige Suchverhalten in der deutschsprachigen Wikipedia untersucht.

In zwei unabhängigen Studien kamen ergänzend explizite (Fragebögen) und implizite (Eyetracking) Methoden zum Einsatz, die ermöglichten, die subjektiven Bewertungen der Inhalte durch den Nutzer zu protokollieren. In beiden Studien konnte gezeigt werden, dass die Suchsituation, hier vor allem im Sinne des Grads der Definiertheit und der Endlichkeit des Suchziels, sowohl Suchstrategie und Informationsrezeption beeinflussen, als auch die Wünsche der Suchenden nach bestimmten Inhaltsstrukturen. Weiter bestätigten die Untersuchungen, dass die Zufriedenheit der Nutzer nicht nur von einem inhaltlichen Sucherfolg abhängt, sondern auch stark mit einer jeweils angemessenen Informationspräsentation korreliert. In Zusammenfassung der eigenen Studien und der Ergebnisse, die der aktuellen Forschungsliteratur entnommen wurden, kam diese Arbeit zu dem Schluss, dass die Situation des Nutzers, bestehend aus einer Vielzahl an Komponenten, an verschiedenen Stellen einen messbaren Einfluss auf die Informationsrezeption des Nutzers beim Information Retrieval in der deutschsprachigen Wikipedia hat.



## ***Abstract***

This doctoral thesis contributes to an understanding of situational perception of information in the Information Retrieval Process in general and in particular searching German Wikipedia. This study advances former research which showed that due to the increasing amount of available digital information in the World Wide Web, the distinct evaluation of search results still leaves a tremendous amount of workload for the user to select currently relevant information.

The research of this thesis uses two approaches: In a theoretical approach the complexity of information seeking tasks and the different situations in which search tasks are occurring are pictured as well as the possible influences of these situations on how users assign relevance to information. Based on those findings the situational meaning of information when the German Wikipedia is searched was investigated with two empirical studies.

For a better insight of the information seeking behavior depending on the user's particular task, different methods for explicit (questionnaires) and implicit (eye tracking) feedback were used. Both studies indicated that the search task influences information seeking behavior and the most preferred content elements. Especially the degree of precision and finiteness of a search goal determines information seeking strategies, which also - using the user actions as a feature - can be distinguished automatically. Along with that the results suggest that an appropriate manner of the information presentation matters as much as the success with regard to the content of a search. This work also describes all individual parameters which play a role for the user's individual and situational information perception.

With respect to the theoretical and the empirical analyses the results of this thesis indicate that the information seeking process for searching German Wikipedia is partially strongly influenced by the user's context.

# Inhalt

<b>1</b>	<b>Einleitung</b>	<b>13</b>
<b>2</b>	<b>Agenda</b>	<b>19</b>
2.1	Unterschiedliches Informationsverhalten	19
2.1.1	Weitere Einflüsse im Kontext einer Informationssuche	20
2.1.2	Modellhafte Darstellung der Suche und des Kontextes	21
2.2	Adaptivität für höhere Nutzerzufriedenheit	21
2.3	Empirische Untersuchungen	22
2.3.1	Inhaltsverteilung in der deutschsprachigen Wikipedia	23
2.3.2	Präferenzen für Inhaltselemente	23
2.4	Struktur der Arbeit	24
<b>3</b>	<b>Motivation: individuelle Suche und Vorstellung des Nutzers</b>	<b>27</b>
<b>4</b>	<b>Abgrenzung des Forschungsgebiets</b>	<b>35</b>
4.1	Kontext in verschiedenen Forschungsbereichen der Informationswissenschaft	35
4.2	Implementierung affektiver und kontextsensitiver Interfaces	36
4.3	Beschränkung auf eine Auswahl an Einflussgrößen	37
4.4	Datenerhebung anhand der deutschsprachigen Wikipedia	37
<b>5</b>	<b>Unterschiedliche Nutzungsszenarien</b>	<b>41</b>
5.1	Suche als Teil unseres Alltags	41
5.2	Information-Seeking und Information Behaviour	43
5.3	<i>Searching for a reason</i> – Warum sucht der Mensch?	52
5.4	Unterscheidung nach dem Suchziel	55
5.5	Unterscheidung nach der Motivation	60
	<i>Non-work-based Szenarien - die casual leisure-Theorie</i>	62
5.6	Unterschiedliche Szenarien – weitere Einflüsse	65
5.6.1	Durchsuchter Wissensbereich	65
5.6.2	Art des Inhaltstyps – Betonung des Medientyps Bild	66
5.6.3	Erfahrung und Vorkenntnis	66
5.6.4	Persönliche Faktoren	67
5.6.5	Der Lerntyp	67
5.6.6	Schwierigkeit und Komplexität des Tasks	68
5.6.7	Einfluss von Emotionen	68
<b>6</b>	<b>Präzisierung in einem Modell</b>	<b>73</b>
6.1	Entwicklung eines Modells	73
6.2	Das Modell	76
6.2.1	Der Nutzer und sein Task	77
6.2.2	Der weitere Nutzerkontext	82

6.2.3	Taskdurchführung und Suchverhalten .....	83
<b>7</b>	<b>Berücksichtigung der Nutzungsszenarien .....</b>	<b>87</b>
7.1	<i>Cognitive Viewpoint</i> und <i>Situational Relevance</i> .....	87
7.1.1	Information als subjektives Konzept .....	88
7.1.2	<i>Cognitive Viewpoint</i> .....	90
7.1.3	Situative Relevanz.....	93
7.2	Zufriedenheit des Nutzers als oberste Priorität .....	95
7.2.1	Adaptive Mensch-Maschine-Interaktion.....	98
7.2.2	Adaption im Information Retrieval.....	99
7.2.3	Adaption mittels Berücksichtigung der Nutzerinteraktion.....	103
<b>8</b>	<b>Ergebnisse der Literatur und offene Fragestellungen .....</b>	<b>109</b>
<b>9</b>	<b>Nutzungsszenarien in der Wikipedia .....</b>	<b>111</b>
9.1	<i>Work-based</i> Nutzung der Wikipedia .....	111
9.2	Freizeitssuche in der Wikipedia .....	111
9.3	Geschichte und Entwicklung der Wikipedia .....	112
9.4	Inhaltstiefe, Details und Sprache in der Wikipedia.....	113
9.5	Inhaltsstrukturen .....	114
9.6	Informationsqualität in der Wikipedia.....	116
9.7	Verteilung der Inhalte auf die Wissendomänen.....	119
9.7.1	Bestehende Ordnung in der Wikipedia.....	121
9.7.2	Anforderungen an das Kategoriensystem.....	124
9.7.3	Struktur des geplanten Ordnungssystems .....	125
9.7.4	Das Kategoriensystem und seine Zuweisungsregeln.....	126
9.7.5	Untersuchungen der Inhaltsverteilung mit Hilfe des Kategoriensystems.....	129
<b>10</b>	<b>Bilder bei der Suche in der Wikipedia.....</b>	<b>137</b>
10.1	Bilder in der Wikipedia.....	139
10.2	Der Begriff „Bild“ .....	141
10.3	Verwendung von Bildern.....	142
10.4	Individueller, situationsabhängiger Nutzen von Bildern.....	145
10.5	Situative Relevanz von Bildern .....	146
10.6	Gezielte Suche und Klassifikation von Bildern.....	146
10.7	Funktionen von Bildern.....	149
10.8	Studie zur Verwendung und Auswahl von Bildern .....	150
10.8.1	Aufbau.....	151
10.8.2	Auswertung .....	152
10.8.3	Folgerungen aus der Studie.....	156
10.9	Anforderungen an ein Ordnungssystem für Bilder .....	157
10.10	Fazit .....	157
<b>11</b>	<b>Nutzer-Präferenzen in Abhängigkeit vom Nutzungsszenario .....</b>	<b>159</b>

11.1	Aufbau .....	162
11.1.1	Methodenauswahl – Emotionen .....	162
11.1.2	Definition und Begrifflichkeit von Emotion – Modelle.....	162
11.1.3	Subjektive Methoden um Emotionen zu erfassen .....	171
11.1.4	Objektive Arten, Emotionen zu erfassen .....	176
11.1.5	Methodenauswahl – Nutzerinteraktion .....	180
11.1.6	Grundlagen zum Aufbau des Experiments.....	184
11.1.7	Verwendung des Modells im Experiment .....	186
11.2	Erste Datenerhebung.....	190
11.2.1	Aufbau der Studie.....	190
11.2.2	Die Aufgaben .....	191
11.2.3	Formulierung der Aufgaben .....	192
11.3	Auswertung der Daten .....	196
11.3.1	Demographische Angaben.....	197
11.3.2	Verarbeitung der Videos des Eyetrackers.....	198
11.3.3	Kruskal-Wallis-Test zur Überprüfung von Unterschieden.....	199
11.3.4	Vergleich der Mittelwerte nach Aufgabentypen .....	204
11.3.5	Verteilung der Wahrscheinlichkeiten .....	209
11.3.6	Analyse der Verteilungen.....	212
11.3.7	Automatische Klassifikation der Daten - taskweise .....	219
11.3.8	Weitere Einflussfaktoren.....	224
11.3.9	Regressionsanalyse .....	230
11.4	Wahl der Methoden .....	234
11.4.1	Eyetracker und Fragebögen .....	234
11.4.2	Elektromyograph.....	235
11.4.3	Self Assessment Manikin.....	236
11.4.4	Browserlogs .....	237
12	Zweite Studie zum Kontextabhängigen Nutzerverhalten.....	239
12.1	Aufbau und Ablauf .....	239
12.2	Verarbeitung der Daten .....	240
12.3	Auswertung der Daten .....	245
12.3.1	Demografische Angaben .....	245
12.3.2	Auswertung des emotionalen Zustandes der Probanden nach jedem Task .....	246
12.3.3	Präferenz für Inhaltselemente im Zusammenhang mit dem Tasktyp .....	251
12.3.4	Auffinden der gesuchten Information .....	253
12.3.5	Angaben zum Suchergebnis und Suchverlauf.....	255
12.4	Auswertung der Videolabels .....	257
12.4.1	Korrelationen Videolabels .....	264
12.4.2	Zusammenfassung der Auswertung .....	265

12.5	Weitere Arbeiten.....	265
12.5.1	Zusammenhang zwischen durchschnittlicher Sakkadenlänge und Task-Typ .....	266
12.5.2	Einflussfaktor thematisches Vorwissen und Erfahrung mit der Wikipedia .....	267
12.5.3	Gelesene Textmenge in Abhängigkeit vom Tasktyp .....	268
12.6	Einfluss der Domain auf die Präferenzen für Inhaltselemente .....	268
12.6.1	Aufbau der Studie.....	270
12.6.2	Ergebnisse.....	271
13	Auswertung des Modells .....	277
14	Diskussion der Ergebnisse .....	279
15	Limitierungen der Arbeit .....	289
16	Fazit.....	293
17	Ausblick .....	297
17.1	Berechnung der Übergangswahrscheinlichkeiten.....	297
17.2	Weitere Aspekte zur adaptiven Inhaltspräsentation.....	300
17.3	Fehlende Daten im Kontextmodell .....	301
17.4	Evaluation der Erkenntnisse dieser Arbeit.....	303
17.5	Methodische Verbesserungen.....	303
17.5.1	Affektive Aktionen und Reaktionen.....	303
17.5.2	Datenerhebung in Feldexperimenten.....	304
17.5.3	Berücksichtigung von Kausalitäten .....	304
17.5.4	Weitere Untersuchungsgegenstände .....	304
18	Verzeichnisse .....	305
18.1	Abbildungsverzeichnis.....	305
18.2	Tabellenverzeichnis .....	307
18.3	Abkürzungsverzeichnis.....	312
19	Literaturverzeichnis .....	313
20	Anhang .....	343
1.	Klassifikation für Bilder mit [Schubart, 2012] .....	343
2.	Ergebnisse der Mozartstudie .....	346
3.	Tasks aus Eyetrackerstudie eins: .....	347
4.	Fragebögen der Studie eins im SS 2012:.....	348
4.1.	Pre Study (Testperson# ).....	348
4.2.	Pre Task (Task# ), (Testperson# ).....	350
4.3.	Post Task (Task # ), Testperson (# ).....	351
4.4.	Post Study (Testperson# ).....	352
5.	Daten aus Studie eins.....	353
5.1.	Verteilung von Interesse und Themenvorkenntnis, alle Aufgaben außer B .....	353
5.2.	Beispiel Videolabels aus Anvil, extrahiert .....	354
5.3.	Häufigkeiten und Lageparameter absolute Häufigkeiten Studie eins .....	355

5.4.	Test auf Normalverteilung der Videolabelkombinationen.....	359
5.5.	Test auf Varianzhomogenität der drei Aufgabengruppen.....	360
5.6.	Deskriptive Statistik der Häufigkeiten .....	361
5.7.	Einfluss der emotionalen Lage auf die Verwendung der Inhaltselemente .....	363
5.8.	Kruskal- Wallis Test über die ersten 50 frames.....	366
5.9.	Einfluss des Geschlechts auf das Suchverhalten .....	367
5.10.	Einfluss des Vorwissens auf die Verwendung der Inhaltselemente .....	368
5.11.	Korrelation Vorwissen und Interesse am Aufgabenthema.....	369
5.12.	Korrelation der Videolabels .....	369
6.	Aufgaben aus Studie zwei .....	371
7.	Fragebögen aus Studie zwei.....	372
7.1.	Pre Study Fragebogen.....	372
7.2.	Pre Task Fragebogen .....	373
7.3.	Post Task Fragebogen.....	373
8.	Daten aus Studie zwei.....	375
8.1.	Lageparameter und Test auf Normalverteilung den Angaben FB.....	375
8.2.	Deskriptive Statistik der absoluten Häufigkeiten .....	376
8.3.	Test auf Normalverteilung.....	386
8.4.	Test auf Varianzhomogenität.....	388
8.5.	Relative Werte .....	388
8.6.	Themenvergleich L1 und L2.....	390
8.7.	Vergleich der Gruppen emotional positiv und negativ .....	391
8.8.	Vergleich der Gruppe viel und wenig Wikipedia -Vorkenntnis .....	393
8.9.	Vergleich der Gruppen viel und wenig Themen-Vorkenntnis .....	394
8.10.	Korrelation der Videolabels .....	397







## 1 Einleitung

*„Weiß jemand, ob eine Katze auch Milchzähne hat?“*

*„Welche der Liparischen Inseln ist denn eigentlich am schönsten, um dort Urlaub zu machen?“*

Wir sind daran gewöhnt, dass Möglichkeiten zur Informationssuche immer zur Verfügung stehen. Lexikalische Werke und Bibliotheken zu schaffen, um die Verfügbarkeit von Wissen und Information zu erhöhen, hat eine lange Tradition. Das Ziel der Autoren der Lexika in Antike und Mittelalter war es, das Wissen ihrer Zeit oder das eines Sachgebietes vollständig zu sammeln und zu systematisieren, um es so dem Leser leichter zugänglich zu machen [Lenz, 1990].

Diese Vereinfachung des Zugangs zu Wissen und Information ist aber nicht mehr vergleichbar mit dem Zustand heute, in dem eine Bibliothek auf dem mobilen Endgerät oder eine freie Enzyklopädie im Internet zur Verfügung stehen.

Die Suche nach Informationen in digitalen Systemen ist damit ein Teil unseres Alltags geworden [Lindley et al., 2012; Morita und Shinoda, 1994]. Aufgrund der Menge und Multimedialität der verfügbaren Informationen treten allerdings neue Probleme auf, die die vorliegende Arbeit wesentlich motiviert haben.

Diese neuen Probleme, vor allem das Überangebot an Informationen, sind dergestalt, dass alleine die Verfügbarkeit und der Zugang zu Wissen und Information im Falle des Information Retrievals nicht immer ausreichen, um den Nutzer tatsächlich situativ in angemessenem Maße zufrieden zu stellen [Ingwersen und Järvelin, 2005; Marchionini, 2006].

Die Ansprüche der Nutzer sind hoch; ausführliche Suchanfragen, gegebenenfalls unter Verwendung logischer Operatoren, oder lange Recherchearbeiten werden nicht akzeptiert. Auch bei ausreichend vorhandener Informationskompetenz dauert die Selektion der tatsächlich gewünschten Inhalte unverhältnismäßig lange [Morita und Shinoda, 1994], ist viel Kenntnis über das verwendete System und/oder Thema notwendig [Brand-Gruwe et al., 2005] oder sie führt sogar zu einer so großen Frustration des Nut-

zers, so dass er eine Suche erfolglos beendet, obwohl die benötigte Information vorhanden gewesen wäre [Wu und Wu, 2011].

Dieser Auswahl- und Suchprozess nach den situativ, also in diesem Moment relevanten Informationen wird dabei von der Vorstellung des Nutzers geleitet [Ingwersen und Järvelin, 2005; Wang und Soergel, 1998]. Diese Vorstellungen wiederum sind von Nutzer zu Nutzer unterschiedlich, da auch das Vorgehen bei der Informationssuche nicht einheitlich ist und vom jeweiligen Nutzungsszenario abhängt [Wildemuth und Freund, 2009]. Die Handlungsrelevanz der angebotenen Informationen wird dabei von verschiedenen Faktoren – z.B. was ist bereits bekannt, was ist die Aufgabe des Nutzers etc. [Wilson, 1999] – beeinflusst. Das bedeutet in der Konsequenz, dass der Nutzer bei seiner Suche bzw. der Selektion der angebotenen Inhalte nur dann unterstützt werden kann, wenn etwas über seine Kriterien zur Auswahl bzw. die Situation und Aufgabe, aus der diese abgeleitet werden können, bekannt ist.

Problematisch dabei ist, dass dieser Informationsbedarf, wenn er überhaupt konkret in der Vorstellung des Nutzers existiert, mit der Eingabe eines Stichwortes zur Suche nicht mit transportiert wird [White und Kelly, 2006]. Das System beantwortet diese Eingabe im Suchfeld aus Systemsicht [Kuhlthau, 1991], diese muss aber nicht zwingend der Sichtweise (des Problems) des Nutzers entsprechen, vor allem, wenn dieser (noch) keine konkrete Vorstellung seines Informationsbedarfs hat sondern sich nur eines *anamolous state of knowledge* [Belkin, 1980] bewusst ist. Gerade hierbei würde die Kenntnis der den Nutzer umgebenden Faktoren wie seine Sachkenntnis helfen, den Informationsbedarf zu konkretisieren.

Der semantische Abgleich von Suchtermen mit den Inhalten einer Wissensquelle erfolgt dabei nach den probaten Regeln des Information Retrievals und das System liefert zunächst alle im klassischen Sinne des Information Retrievals als relevant bewerteten Dokumente zurück. Bei der Wikipedia ist dies i.d.R. ein Artikel oder – sollte zu den eingegeben Suchtermen kein Artikel bestehen – die Ergebnisse einer Volltextsuche. Die situationsgebundene Auswahl der handlungsrelevanten Informationen erfolgt anschließend durch den Nutzer.

Die Internet-Enzyklopädie Wikipedia erfreut sich einer großen Beliebtheit und rangiert derzeit<sup>1</sup> auf Platz sechs der weltweit meistbesuchten Websites [Alexa, 2013]. Mehr als 36% der über 18-jährigen Amerikaner nutzen die Wikipedia [Rainie und Tancer, 2007]. Sie zeichnet sich durch ihren Umfang, ihre Verfügbarkeit in vielen Sprachen und die Bandbreite ihrer Inhalte aus [Head und Eisenberg, 2010]. Viele verschiedene Inhaltstypen wie Bilder, Listen oder Texte erwarten dort den Nutzer [Clark et al., 2012]. Die Artikel in der Wikipedia, vor allem die als exzellent oder lesenswert ausgezeichneten, sind teilweise sehr lang. Der Nutzer ist hier nicht mehr in der Lage, alles zu überblicken, um schnell eine eigene, der Situation entsprechende Auswahl an Inhalten zu treffen. An diesem Punkt ist es deshalb vorstellbar, dass eine Unterstützung des Systems z.B. bei der Auswahl der situativ relevanten Inhalte, hilfreich wäre.

Die Wikipedia kann nachvollziehbar für verschiedene Anliegen der Nutzer herangezogen werden. Man kann dort Fakten nachschlagen (z.B. die Frage nach der Höhe des Matterhorns), sich aber auch ausführlich über ein Thema informieren (zum Beispiel geologische und geographische Eigenschaften der Westalpen und deren Eignung als Urlaubsort etc.). Strukturell berücksichtigt die Wikipedia das zwar bereits durch lexicographische Konventionen bei der Darstellung und mittels Strukturelementen, wie z.B. Übersichtstexten oder einem Inhaltsverzeichnis, allerdings entbindet das den Nutzer nicht davon, diese Strukturelemente auf die jeweilige situative Eignung hin zu evaluieren (je nach Fragestellung).

Eine wesentliche Forschungsfrage dieser Arbeit bezieht sich deshalb auf die unterschiedlichen Nutzungsmöglichkeiten der Wikipedia und weiter, ob die Präferenzen der Nutzer für bestimmte Struktur- respektive Inhaltselemente bzw. die Rezeption der angebotenen Information abhängig von der Nutzungssituation unterschiedlich ausfallen. Damit steht der individuelle und situationsgebundene Selektionsprozess der Inhalte, der nach der Auswahl durch das System entsprechend den Suchtermen in der Wikipedia erfolgt, im Zentrum dieser Arbeit. Ziel ist es, aus diesen individuellen Beobachtungen allgemeine Szenarien abzuleiten und für diese Aufschluss über die jeweilige Informationsrezeption zu erhalten. Die entscheidende Vermutung dabei ist, dass

---

<sup>1</sup>September 2013

<sup>2</sup> Kontext ist in diesem Fall keine fachliche Definition sondern steht für die Situation des Nutzers, die in

die Informationsauswahl und -rezeption abhängig von der Aufgabe des Nutzers und seinen individuellen Eigenschaften (wie Motivation, thematische Vorkenntnis etc.) ist.

Eine Berücksichtigung dieses individuellen Nutzungskontxtes beim der Informationssuche – wie auch bei anderen adaptiven Systemen (z.B. [Kostatinova et al., 2012]) – hat bereits grundsätzlich das Potential gezeigt, die Informationsrecherche in digitalen Informationssystemen zu verbessern [Benyon et al., 1994]. Vor allem die Zufriedenheit des Nutzers mit dem Suchverlauf und den Suchergebnissen werden dabei in den Vordergrund gestellt. Durch maßgeschneiderte, der konkreten Situation angepasste Information, die die Umstände der Interaktion und die daraus entstehenden Präferenzen des Nutzers berücksichtigt, kann die Suche effektiver und für den Nutzer angenehmer gestaltet werden [Belkin, 2006].

Um aber den Gedanken an eine Unterstützung des Nutzers durch eine Vorauswahl von Information durch das System realistisch verfolgen zu können, muss bekannt sein, welche Präferenzen der Nutzer in welcher Situation hat und welche unterscheidbaren Situationen auftreten können. Es müssen Kriterien gefunden werden, nach denen das System eine Vorauswahl der angebotenen Informationen treffen kann. Gleichzeitig muss aber auch gegeben sein, dass die Nutzer tatsächlich ein unterschiedliches Rezeptionsverhalten (in Form von der Bevorzugung unterschiedlicher Inhalte zur Problemlösung) in Abhängigkeit ihrer jeweiligen Aufgabe und Situation haben. Diese Unterschiede müssen weiter systematisierbar sein und in einem Zusammenhang mit der jeweiligen Nutzungssituation stehen.

Diese Frage ist hinsichtlich der Nutzung der Wikipedia noch offen. Da aber gerade die Wikipedia eine sehr häufig benutzte Informationsquelle ist und zu unterschiedlichen Zwecken verwendet werden kann [Lim, 2009], wird dieser Fragestellung in dieser Arbeit nachgegangen. Dabei geht es nicht darum, dass besonders die Wikipedia von einer adaptiven Informationspräsentation profitieren würde. Sie bietet aber als thematisch vielfältiges und geschlossenes System ein ausreichendes aber dennoch gut abgegrenztes Forschungsfeld für Untersuchungen zur situationsabhängigen Rezeption und Verwendung von Information beim Information Retrieval.

### *Hauptziele der Arbeit*

In dieser Arbeit wird nach Hinweisen geforscht, die zeigen, dass erstens Nutzungssituationen im Information Retrieval unterscheidbar und beschreibbar sind, dass zweitens auch die Nutzungsszenarien der deutschsprachigen Wikipedia entsprechend der für diese Arbeit relevanten Aspekte unterscheidbar sind und dass drittens eine Berücksichtigung des jeweiligen Nutzungskontexts<sup>2</sup> das Potential birgt, die Zufriedenheit der Nutzer und die Qualität der Suchergebnisse zu verbessern. Es werden somit drei Hypothesen formuliert:

1. Die Nutzungsszenarien der Wikipedia sind heterogen. Aufgabe und generelle Ausgangssituation variieren zwischen den Nutzern entsprechend den generellen Unterscheidungen von Aufgabentypen im Information Retrieval.
2. Eine Berücksichtigung des Nutzerkontexts im Information Retrieval verbessert die Nutzerzufriedenheit.
3. Die Präferenzen für die Inhaltselemente der Wikipedia sind in den unterschiedlichen Nutzungsszenarien *nicht* gleich.

Diese Hypothesen werden in der Arbeit motiviert und sowohl mit einem theoretischen als auch einem empirischen Ansatz untersucht. Das generelle Ziel dieser Arbeit sind Erkenntnisse, wie die Erforschung und Berücksichtigung der jeweiligen Nutzungssituation im Information Retrieval am Beispiel der deutschsprachigen Wikipedia konkret aussehen. Ein Fokus liegt dabei darauf, von der vorhandenen Theorie zur praktischen Untersuchung zu kommen.

---

<sup>2</sup> Kontext ist in diesem Fall keine fachliche Definition sondern steht für die Situation des Nutzers, die in Kapitel 6 präzisiert wird. Dies dient als Realdefinition von Kontext bzw. Situation für den Rahmen dieser Arbeit.



## 2 Agenda

Im Folgenden werden die Schritte dieser Arbeit kurz skizziert.

### 2.1 Unterschiedliches Informationsverhalten

Die Beobachtung und Untersuchung des Suchverhaltens im Information Retrieval beschäftigt die Forschung umso mehr, seit die Informationsmenge, die im World Wide Web zur Verfügung steht, rasant ansteigt [Marchionini, 2006].

Eine wesentliche Motivation für diese Arbeit liegt darin, dass die Suche in digitalen Informationssystemen ein Teil unseres Alltags geworden ist [Johnson, 2002; Lindley et al., 2012]. Deshalb wird hier dargestellt, dass dadurch auch die Bandbreite an Nutzungsszenarien und damit wiederum die Anforderungen und vor allem das Verhalten der Nutzer vielfältiger und heterogener geworden ist und weiter wird [siehe u.a. Sohn et al., 2008; Wildemuth et al., 2010; Belkin, 2006; Marchionini, 2006; Rose und Levinson, 2004].

Heterogenität bedeutet dabei, dass sowohl Ort, Gerät und Umstände der Nutzung als auch die zu lösenden Aufgaben und die Suchauslöser bzw. Motivation zur Suche in zahlreichen Variationen vorkommen.

Zur Verdeutlichung dieser Heterogenität vor allem hinsichtlich der unterschiedlichen Suchsituationen und Motivationen werden die dazu in der Forschungsliteratur vorhandenen Theorien und empirischen Ansätze besprochen. Verschiedene Autoren [z.B. [Marchionini, 2006; Marchionini, 1995; Kang und Kim, 2003; Broder, 2002] zeigen, dass es deutlich unterscheidbare Typen von Suchaufgaben gibt. Sie differenzieren zum Beispiel hinsichtlich der Motivation zur Suche, des Gegenstands, der Ausführenden, des verwendeten Systems, der Vorgehensweise, des (inhaltlichen) Ziels und weiterer Kriterien (siehe dazu [Wildemuth et al., 2010]).

Bei der Inspektion der aktuellen Forschungsliteratur zu der Frage nach Nutzungsszenarien wird weiterhin deutlich, dass die Literatur zwar präzise ist hinsichtlich der Existenz verschiedener Tasktypen (vgl. [Broder, 2002; Marchionini, 2006; Elsweiler und Wilson, 2010]), wobei daraus allerdings nicht eindeutig hervorgeht, ob dieser Umstand auch auf die Wikipedia zutrifft.

Verschiedene Studien [Spoerri, 2007; Lim, 2009] zeigen, dass die Wikipedia regelmäßig als Nachschlagewerk für sog. *work-based* [Borlund, 2000] Szenarien verwendet wird, auch im akademische Kontext [Head und Eisenberg, 2010].

Für diese Arbeit gilt die Annahme, dass eine Bandbreite an Nutzungsszenarien bei der Wikipedia dann gegeben ist, wenn sie auch Gegenstand von sog. *casual-leisure searches* [Elsweiler et al., 2011], also der Suche als reinem Selbstzweck in Freizeitsituationen ist. Deshalb wird, bevor der Frage nach situationsbezogenen Relevanzbewertungen von Inhalten weiter nachgegangen wird, geklärt, ob es plausibel ist, auch für die Wikipedia von *non-work-based* Nutzungsszenarien auszugehen.

### 2.1.1 Weitere Einflüsse im Kontext einer Informationssuche

Die Existenz unterschiedlicher Tasktypen erklärt teilweise verschiedenes Nutzerverhalten [Marchionini, 2006]. Darüber hinaus wird hier noch die Frage nach weiteren Einflüssen, die das Vorgehen, die Ziele und die subjektive Relevanz von Inhalten beim Information Retrieval beeinflussen, besprochen. Auf Basis der Forschungsliteratur werden deshalb Faktoren identifiziert, die potentiell Vorgehen und Präferenzen der Nutzer beim Information Retrieval steuern.

Aus zwei Gründen wird in dieser Arbeit neben der kognitiven (Aufgabe, Vorwissen etc...) auch die *affektive* Ebene beim Information Retrieval [Nahl und Bilal, 2007] berücksichtigt:

1. Emotionen (z.B. genervt, gelangweilt, angespannt, entspannt, glücklich, gestresst etc.) beeinflussen immer unsere Entscheidungen [De Sousa, 1990]. Folglich lassen sich auch während des Suchverlaufs Emotionen und Emotionswechsel nachweisen [Kuhlthau, 1991; Moshfeghi und Jose, 2013].
2. Die emotionale Ebene des Nutzers spielt bei der Freizeit-Suche (sog. *casual-leisure* Szenarien) eine wesentliche Rolle: Steht beim Fakten-Retrieval das Finden bestimmter, relevanter Informationen im Vordergrund, so ist dagegen beim reinen „Freizeitvergnügen“ die affektive Ebene wichtiger. Erklärtes Ziel einer solchen Freizeitsuche kann z.B. sein, die eigene Stimmung zu *verbessern* [Elsweiler und Wilson, 2010]. Somit steht dieser Aspekt gleichwertig neben dem Finden relevanter Fakten.



### 2.1.2 Modellhafte Darstellung der Suche und des Kontextes

Die Literatur zu den verschiedenen Einflussfaktoren auf das Suchverhalten ist umfangreich. Es kommen dort viele verschiedene Ansätze und Methoden zum Einsatz. Um die einzelnen Ergebnisse zueinander in Beziehung setzen zu können wird in dieser Arbeit deshalb eine Formalisierung der genannten Einflussfaktoren und des Nutzerverhaltens in Form eines Modells vorgenommen und um neben Inhaltsziel und Motivation weitere Aspekte einer Suche (z.B. Selektion der Inhalte etc.) darstellen zu können. Dies wird unter Berücksichtigung existierender Modelle aus dem Bereich des *Information Seeking* und des *Information Behaviour*, die den Suchprozess formal wiedergeben (z.B. [Wilson, 1999; Hearst, 2009; Bates, 1989 oder Kuhlthau, 2004]) vorgenommen.

Da es sich bei den vorhandenen Modellen i.d.R. um die Darstellung abstrakter Konzepte handelt (z.B. *Information need*, *information interaction* oder *user satisfaction*) werden diese hier konkretisiert. Das dient dazu, erstens das bisher beschriebene Themengebiet des Kontextes im Information Retrieval (Suchszenarien, weitere Einflüsse, Suchverhalten) zusammen zu fassen und es zweitens hinsichtlich der Suche in der Wikipedia zu präzisieren. Dieses Modell soll vorhandene Modelle (z.B. [Byström und Järvelin, 1995 oder Ingwersen und Järvelin, 2005]) *ergänzen*.

## 2.2 Adaptivität für höhere Nutzerzufriedenheit

Aus der vorgenommenen Darstellung der Existenz unterschiedlicher Nutzungsszenarien leitet sich die anschließende Frage ab, ob es sinnvoll ist, diese bei der Informationspräsentation zu berücksichtigen.

Auf Basis der vorhandenen Forschungsliteratur wird deshalb zunächst dargelegt, dass Adaptivität, also eine Anpassung an bestimmte Nutzereigenschaften, ein probates Mittel ist, um eine höhere Nutzerzufriedenheit zu erreichen [Trumbley et al., 1994]. Es wird weiter gezeigt, dass auch im Bereich des Information Retrieval mit adaptiven Systemen bereits verbesserte Ergebnisse erzielt werden konnten (u.a. [Lau et al., 2008; Crestani und Van Rijsbergen, 1997; Wen et al., 2007]).

Bei der Analyse vorhandener Ansätze im adaptiven Information Retrieval wird allerdings deutlich, dass sich die Forschung auf diesem Gebiet eher auf Systemeigenschaften bezieht. Die Berücksichtigung der genannten *persönlichen* Faktoren wie z.B. dem thematischen Vorwissen oder Sachverständnis ist ein Unterbereich des adaptiven In-

formation Retrieval und gestaltet sich schwieriger als „nur“ die Anpassung von Funktionalitäten oder die semantische Erweiterung von Suchanfragen [Belkin, 2006].

Es wird deshalb deutlich gemacht, dass die Berücksichtigung dieser persönlichen Faktoren ebenfalls sinnvoll ist, wenn eine situationsrelative Informationspräsentation angestrebt wird (z.B. [Fox et al., 2005; Kelly und Fu, 2007]), die die Nutzerzufriedenheit verbessern soll. Um situativ relevanten Informationen über den Nutzer und seine Ziele zu erhalten, kann dessen Interaktion mit den Inhalten herangezogen werden. Es werden deshalb verschiedene Methoden, mit deren Hilfe bereits zur Laufzeit Informationen über Zufriedenheit und Ziele der Nutzer gewonnen werden können, vorgestellt (z.B. [Claypool et al., 2001; Joachims et al., 2007]).

Ein theoretischer Aspekt, der die Plausibilität des Anliegens, die persönlichen situativen Faktoren des Nutzers zu berücksichtigen, bestärkt, wird weiterhin durch Vorstellung der Konzepte der situativen Relevanz [Borlund, 2000] und dem sog. *cognitive viewpoint* [Ingwersen und Järvelin, 2005] gegeben. Diese basieren auf einer holistischen Sichtweise der Informationssuche – als ergänzender Aspekt zum „klassischen“ Information Retrieval. Wesentlich dabei ist, dass diese darlegt, warum auch der persönliche Kontext die Rezeption der Inhalte beeinflusst.

Für eine situationsrelative Informationspräsentation ist dies aber nur dann eine Rechtfertigung, wenn die Frage nach unterschiedlichen Relevanzbewertungen der Inhalte in Abhängigkeit von Nutzungsszenarien bejaht werden kann. Hinweise darauf sind z.B. bei Cutrell und Guan [Cutrell und Guan, 2007] zu finden. Es bleibt offen, wie sich die Präferenzen für bestimmte *Inhaltselemente*<sup>3</sup> in der Wikipedia abhängig vom Nutzungsszenario gestalten.

## 2.3 Empirische Untersuchungen

Den offenen Fragen, die nicht durch Analyse der Forschungslage geklärt werden konnten, wird mit empirischen Ansätzen nachgegangen.

---

<sup>3</sup>Also die tatsächlichen Inhalte die der Nutzer zur Problemlösung heranzieht, nicht die Relevanz allgemeiner Dokumente (das wären hier entsprechend die Wikipedia-Artikel bzw. deren Ranking).

### 2.3.1 Inhaltsverteilung in der deutschsprachigen Wikipedia

Eine grundlegende Frage ist, ob die Wikipedia ebenfalls Gegenstand der *work-based* [Borlund, 2000] *und* der Freizeitsuche ist. Die Literatur zeigt, dass Suche grundsätzlich unterschiedlich motiviert sein kann, nicht aber ob dies konkret bei der Wikipedia der Fall ist.

Um einen Ansatzpunkt für die Annahme, dass die Wikipedia auch aus reinen Freizeit-zwecken zum Surfen und zur explorativen Suche verwendet wird, zu erhalten, wurden die Inhalte der Internetenzyklopädie untersucht.

Dabei wird ein hoher Anteil von Artikeln über Themen wie Musikgruppen, Filme, Sportereignisse und Personen aus dem Sportgeschehen als Hinweis gesehen, dass verstärkt auch eine Beschäftigung mit diesen Themen in der Freizeit (z.B. im Sinne von [Adafre und Rijke, 2006]) vorkommen kann. Diese Themen haben eine höhere allgemeine Wahrscheinlichkeit, Gegenstand einer Freizeitsuche zu sein, weshalb der Anteil der Themen aus diesem Bereich als Indikator für eine mögliche Verwendung der Wikipedia zur Freizeitsuche herangezogen wird. Damit wird vor allem auf eine unterschiedliche *Motivation* zur Suche referiert, nicht auf den Aufgabentyp, der auch in den Freizeitbereichen der vorgenommenen Unterteilung (siehe dazu Kapitel 5.4) entspricht. Da in der Literatur keine Aussagen über die Themenverteilung in der deutschsprachigen Wikipedia unter diesem Aspekt vorhanden sind, wurden die Inhalte mittels einer Artikel-Stichprobe untersucht, ebenso eine Stichprobe aus der Brockhaus Enzyklopädie, um eine mögliche Verschiebung der Themenschwerpunkte zu entdecken.

### 2.3.2 Präferenzen für Inhaltselemente

Eine adaptive Informationspräsentation ist nur dann sinnvoll, wenn die subjektive Bewertung der Inhaltselemente bei der Suche in der Wikipedia tatsächlich kontextrelativ ist. Das wird hier zunächst am Beispiel des Inhaltstyps „Bild“ überprüft. Im Vordergrund steht dabei die Hypothese, dass gerade Bilder situationsgebunden unterschiedlich geeignet sind und dass bei der Auswahl und Bewertung von Bildern auch die subjektive Sicht des Nutzers für die jeweilige Relevanz ausschlaggebend ist [Ingwersen und Järvelin, 2005]. Deshalb wird die subjektive Auswahl und Eignung von Bildern auch in Abhängigkeit ihrer Funktion dem Text gegenüber empirisch untersucht.

Die daran anschließende Frage, ob in den verschiedenen Nutzungsszenarien in der Wikipedia die Präferenzen der Nutzer für die Inhalte unterschiedlich ausfallen, bezieht sich dann auf *alle* in der Wikipedia vorhandenen Inhaltselemente. Dabei wird nicht nur der Einfluss der Aufgabenart, sondern auch der weiterer Kontextfaktoren wie Vorwissen oder die emotionale Lage des Nutzers berücksichtigt. Es wird vor allem auch auf die Problematik der Berücksichtigung des emotionalen Zustands des Nutzers während der Suche eingegangen. In der Forschung stehen dazu verschiedene Methoden zur Auswahl, die jeweils Vor- und Nachteile haben. Diese werden besprochen, um für die Untersuchungen eine geeignete Auswahl treffen zu können.

In den Studien werden die Nutzer mit unterschiedlichen, entsprechend der bisher besprochenen Literatur gestalteten Suchaufgaben konfrontiert, die sie mit Hilfe der Wikipedia zu bearbeiten haben. Ziel der Studien ist es, Hinweise auf die subjektiven Relevanzbewertungen der Suchenden in Abhängigkeit von den unterschiedlichen Aufgaben zu finden.

Da diese Studien zeigen, dass sich die Annahme der nicht einheitlichen Präferenzen bestätigt, werden anschließend noch weitere Überlegungen zu Fragen, die bei einer adaptiven Informationspräsentation zu berücksichtigen sind, angestellt.

Abschließend werden die Erkenntnisse aus den theoretischen und praktischen Ansätzen dieser Arbeit zusammengefasst und besprochen.

## 2.4 Struktur der Arbeit

Entsprechend dem beschriebenen Forschungsdesiderat ist die Arbeit aufgebaut wie folgt:

Nach einer Einleitung und Motivation der Arbeit (Kapitel 3) wird das Forschungsgebiet eingegrenzt (Kapitel 4). Im darauf folgenden theoretischen Teil wird zunächst gezeigt, dass unterschiedliche Nutzungsszenarien von Informationssystemen vorhanden sind, auch, weil die Informationssuche Teil unseres Alltags geworden ist (Kapitel 5.1). Die Modelle des *Information Seeking* und *Information Behaviour* erklären, an welcher Stelle Einflüsse der Nutzungssituation auf die subjektive Relevanz der Inhalte für den Nutzer entstehen (Kapitel 5.2). Die identifizierten Einflussfaktoren werden in Kapitel 5.1 bis 5.6.7 besprochen, der besondere Augenmerk liegt dabei auf den unterschiedli-

chen Motivationen und den inhaltlichen Zielen, die bei einer Informationssuche auftreten können.

Um eine Übersicht zu erhalten und die Ergebnisse aus dem aktuellen Stand der Forschung auf die hier vorliegende Fragestellung zu beziehen werden in Kapitel 5.2 die vorhandenen Ergebnisse in einem Modell zusammengefasst.

Die daran anschließenden Überlegungen gelten der Frage, ob Adaptivität generell und speziell im Information Retrieval sinnvoll ist. Dabei steht vor allem die Adaptivität hinsichtlich persönlicher Faktoren im Gegensatz zur semantischen Adaptivität der Inhalte und/oder der Query im Fokus (ab Kapitel 7).

Anschließend werden die noch offenen Fragestellungen empirisch untersucht. Dazu wird erstens der Frage nach der inhaltlichen Verteilung der Wikipedia-Artikel auf die unterschiedlichen Wissensdomains nachgegangen (ab Kapitel 9.2), zweitens wird ein empirischer Ansatz gewählt, um die subjektive Auswahl von Bildern zu untersuchen (Kapitel 10) und drittens werden zwei Studien durchgeführt, die auf die Fragestellung der generellen Nutzungsszenario-abhängigen Präferenz von Inhaltselementen bei der Informationssuche in der Wikipedia abzielen (ab Kapitel 11). Dazu ergänzende Arbeiten und Ergebnisse werden in den Kapiteln 12.5 und 12.6 besprochen.

Weitere Aspekte, die bei der Berücksichtigung einer situativ angepassten Informationspräsentation zu berücksichtigen sind, werden in den Kapiteln 17.1 und 17.2 aufgegriffen.

Die Ergebnisse werden ab Kapitel 14 besprochen so wie ein Ausblick gegeben auf verbleibende bzw. neue Fragestellungen, die sich aus den Ergebnissen dieser Arbeit ergeben.



### 3 Motivation: individuelle Suche und Vorstellung des Nutzers

Die Relevanz der Fragestellungen dieser Arbeit begründet sich in dem Wunsch, die Informationssuche für den Nutzer zu erleichtern. Dieser Wunsch und die dazugehörigen Fragen kamen in der Forschung zum Information Retrieval bereits an anderer Stelle zum Ausdruck:

Nicholas Belkin, 1980:

*“[...] my view of the function of an IR System: to help people to solve problems, rather than directly to solve problems posed to them [...]”*

[Belkin, 1980, S. 134]

Gary Marchionini, 1989:

*„3. What are the relationships among user, task and search pattern?“*

[Marchionini, 1989, S.3]

Carol Kuhlthau, 1991:

*“The individual is actively involved in finding meaning that fits in in what he or she already knows [...]”*

[Kuhlthau, 1991, S.1]

Gary Marchionini, 2000:

*„People are interested in accomplishing those tasks rather than executing searches“*

[Marchionini et al., 2000, S.1]

Diese Bemerkungen zu dem Zusammenhang zwischen dem Nutzer, seiner Aufgabe und seiner Vorstellung der Zielerreichung werden hier aufgegriffen und auf die konkrete Anwendung der Wikipedia adaptiert. Dabei werden zusätzlich die aktuellen Entwicklungen im Gebiet des IR, wie die Suche als Alltagsbeschäftigung, die universellen Verfügbarkeit von digitalen Inhalten wie der online-Enzyklopädie Wikipedia, sowie deren speziellen Eigenschaften als kollaborative Kreation des Web 2.0 berücksichtigt.

Alle genannten Zitate haben eine Gemeinsamkeit: Die Individualität des „Problems“ einer Person, die subjektiven Lösungsvorstellungen und die davon beeinflusste Relevanz der Information und die Vorgehensweise bei der Suche sind wichtige Aspekte, die beim Information Retrieval berücksichtigt werden sollten.

Zunächst ist „Suche“ beim allgemeinen Information Retrieval definiert als eine Problematik, die sich vor allem mit dem Abgleich von Such- und Anfragetermen so wie dem daraus resultierenden Ranking der Suchergebnisse beschäftigt [Belkin, 1993].

Eine Grundannahme ist dabei, dass ein Dokument umso relevanter ist, je höher die Anzahl der vom Nutzer verwendeten Suchterme in diesem Dokument ist. Diese Relevanz wird nach verschiedenen Algorithmen vergeben [Lewandowski, 2005]. Die Anfrage (Eingabe einer *query*) des Nutzers wird vom System verarbeitet und eine entsprechende Treffermenge - mit oder ohne Ranking der Treffer - zurückgeliefert (unter Berücksichtigung bestimmter Aspekte zur Gewichtung der Terme).

Aber gerade bei der Informationssuche im World Wide Web beschreibt das die Problematik nicht mehr vollständig, da die enorme Menge an verfügbaren Inhalten neue Herausforderungen geschaffen hat [Fonseca et al., 2005]. Auch das Ranking der Trefferdokumente entsprechenden den Algorithmen des klassischen IR ist nicht mehr ausreichend weil erstens die so entstandenen Listen gerade bei der Websuche sehr umfangreich sind (und damit oft nicht komplett gesichtet werden können) und zweitens zusätzliche Informationen über das Suchbedürfnis des Nutzers in der Regel nicht berücksichtigt werden (sowohl inhaltlich im Sinne einer semantischen Anreicherung der Querys als auch kognitiv und affektiv, siehe z.B. [Morita und Shinova, 1994; Cao et al., 2009; Fonseca et al., 2005; White und Kelly, 2006; He et al., 2009; Guo et al., 2008 u.a.]).

Bei der Wikipedia wird in der Regel mit Hilfe eines Stichwortes gesucht. Der diesem entsprechende Artikel enthält alle zu diesem Zeitpunkt in der Wikipedia vorhandenen Informationen. Nach Präsentation dieses entsprechenden Artikels beginnt dann bei der Suche der Auswahlprozess des Nutzers – er selektiert, welche der angebotenen Informationen in diesem Moment *tatsächlich* relevant für ihn sind.

Diese subjektive Auswahl und damit die Tatsache, dass der rein inhaltliche Abgleich von Suchtermen und vorhandener Dokumentenmenge deshalb nicht immer optimal ist, beschreibt auch Carol Kuhlthau [Kuhlthau, 1991]. Sie postuliert, dass der Nutzer



bei der Informationssuche in seinen Entscheidungen, die er während der Suche trifft, von verschiedenen äußerlichen Faktoren geleitet wird, wie etwa seiner Erfahrung, der verfügbaren Zeit zur Lösung der Aufgabe und den spezifischen Erfordernissen, die aus seinem Suchproblem resultieren. Außerdem stellt auch Kuhlthau fest, dass sowohl die genannten kognitiven, als aber darüber hinaus auch *affektive* Einflüsse das Verhalten der Nutzer steuern. Diese Faktoren des Kontextes sind immer vorhanden, werden aber bei der Informationssuche via einer *query* nicht berücksichtigt. Kuhlthau beschreibt den Suchprozess als mehrdimensional:

„An information search is an process of construction which involves the whole experience of the person, feelings as well as actions“ [Kuhlthau, 1991, S. 2]

Diese holistische Betrachtungsweise des Suchprozesses ist plausibel, es haben auch andere Autoren Einflüsse der *whole experience of the person* [Kuhlthau, 1991, S. 2] bereits gezeigt. Es spielen auch Emotionen [Nahl und Bilal, 2007; Axelrod, 2004; Arapakis et al., 2008; Moshfeghi und Jose, 2013], der durchsuchte Wissensbereich [Yan et al., 2011], verschiedene Inhaltstypen wie Bilder [Nöth, 2008; Weidenmann, 1991], die Erfahrung und Vorkenntnis des Suchenden [White et al., 2009; Debowski, 2001; Brand-Gruwel et al., 2005], das Alter [Bilal und Kirby, 2002] und generell die kognitive Wissensstruktur des Nutzers [Ingwersen und Järvelin, 2005] eine Rolle.

Das bedeutet, dass außerhalb des Systems eine weitere Relevanzbewertung der Informationen durch den Nutzer stattfindet, wobei nicht mehr nur Termfrequenzen, sondern die persönliche Sicht der jeweiligen Person herangezogen wird.

Auch Belkin [Belkin, 1993] berücksichtigt bei seinen Überlegungen zum *Information Retrieval as Information Behaviour* [Belkin, 1993, S.1] den Nutzer als zentrale Komponente des Information Retrieval-Prozesses. Dies gründet er vor allem auf die Tatsache, dass der Mensch nicht nur passiver Empfänger einer Nachricht oder eines Inhalts ist, sondern aktiv sucht. Die Suche ist motiviert und gesteuert durch die Ziele der Nutzer. Wird aber der Nutzer im Standard-Information Retrieval Prozess nicht berücksichtigt, dann können diese persönlichen und individuellen Aspekte auch nicht miteinbezogen werden. Belkin [Belkin, 1993] betont aber, dass es für die Erfüllung eines Informations-

bedürfnisses sinnvoller bzw. der *most appropriate way* [Belkin, 1993, S.3] ist, genau *die* Inhalte auszuwählen, die am besten geeignet sind, um genau dieses zu erfüllen. Genau dieser Gedanke wird in dieser Arbeit am konkreten Beispiel der Wikipedia-Suche aufgegriffen: Die nach dem Standard-IR Verfahren besten Ergebnisse werden durch den Nutzer weiter evaluiert, und zwar hinsichtlich der konkreten Nutzungsbedürfnisse, dem aktuellen *need* des Individuums.

Eine sich hier anschließende Frage ist dabei, was unter dem *information need* des Nutzers zu verstehen ist. Wird dieses durch die verwendeten Suchterme abgebildet? Gibt es eine konkrete Frage oder eine unscharfe Vorstellung von den gesuchten Informationen [Taylor, 1962]? In der Literatur wird der Begriff *information need* unterschiedlich definiert und verstanden. In dieser Arbeit wird deshalb das *information need* formal dargestellt und definiert (siehe v.a. Kapitel 5.4). Wie Wilson [Wilson, 2006] betont, ist eine Berücksichtigung dieser *needs* bei der Evaluation des IR-Prozesses, vor allem zum Zwecke einer Verbesserung der Ergebnisse, ein wichtiger Aspekt.

Theoretisch ist der Mensch zunächst selbst in der Lage, selektiv vorzugehen, um sich die jeweils geeigneten Inhalte eines Wikipedia-Artikels oder der Ergebnisliste einer Websuche herauszufiltern. Entsprechend Dervins *sensemaking* [Dervin, 1992] entscheidet das Individuum jeweils so, dass es für es selbst (in Korrespondenz mit seinen aktuellen Bedürfnissen) passend und sinnvoll ist.

Durch die sehr große Menge an Inhalten im *World Wide Web* besteht aber die Gefahr, dass dieser individuelle *Selektionsprozess* zu lange dauert [Morita und Shinoda, 1994] oder sogar ganz ohne Erfolg bleibt [Wu und Wu, 2011], da die Menge an Informationen zu groß, die Informationskompetenz des Nutzers zu klein oder die Umstände der Suche zu kompliziert sind (z.B. wenig verfügbare Zeit, Ablenkung und Störung von außen etc.).

Da der Nutzer bei der Formulierung einer Suchanfrage generell eher unpräzise und Formulierungen verwendet, sind die Ergebnisse oft nicht entsprechend seinem Informationsbedürfnis [Fonseca et al., 2005] und der Nutzer muss aus der Auswahl an Dokumenten wiederum eine Auswahl treffen [Kelly und Teevan, 2003; Morita und Shinoda, 1994]. Dabei wird er nur die Dokumente oder Dokumententeile berücksichtigen, die seiner Vorstellung entsprechen [ebd.; Hearst, 2009].

Bei der Wikipedia sind die Artikel ebenfalls zu großen Teilen sehr lang und umfangreich. Vor allem bei der Nutzung mobiler Endgeräte, die die teilweise große Menge an Informationen in Form langer Artikel in der Wikipedia aufgrund kleiner Displays schlecht darstellen können, werden die individuell sinnvoll erscheinenden Entscheidungen bei der Inhaltsauswahl allein aus Gründen der Übersichtlichkeit schwierig. Eine generelle Veränderung der Inhalte (z.B. Kürzung der Artikel) ist bei der hier vermuteten stark *heterogener* Nutzung keine optimale Lösung. Daraus entsteht die Frage, ob es sinnvoll und möglich ist, *Unterstützung* durch das System bei den jeweiligen individuellen Auswahl- und Entscheidungsprozessen zu bieten.

In der Wikipedia wird bereits durch bestimmte strukturelle Maßnahmen wie Infoboxen, Einleitungen und ein Inhaltsverzeichnis versucht, dem Nutzer den Einstieg in die Selektion zu erleichtern. Nicht berücksichtigt werden (können) aber bestimmte situationsabhängige Wünsche der Nutzer. Für eine Unterstützung beim individuellen Auswahlprozess durch den Nutzer muss geklärt werden, ob dieser Auswahlprozess situationsabhängig verläuft und wenn ja, welche Zusammenhänge es gibt.

Auch in anderen Bereichen des Information Retrieval ist der Nutzen einer Informationspräsentation, die die aktuelle Situation des Nutzers berücksichtigt, vorstellbar: Informationssysteme für Menschen mit einer Krankheit oder auch Leiden wie Übergewicht, sollen dazu beitragen, im Alltag besser mit ihrer Krankheit umzugehen und entsprechende Informationen bereitstellen [Zini und Ricci, 2012].

Die Nutzung der Systeme zur Unterstützung einer gesundheitsförderlichen Lebensweise ist aber noch nicht optimal. Bei sensiblen Themen ist es unter Umständen schwierig, die richtige *Menge* und *Präsentationsform* der Informationen zu finden, um den Nutzer zu unterstützen und nicht zu „verschrecken“. Eine Studie von Spink und Cole [Spink und Cole, 2001] zeigte, dass krebskranke Menschen vermehrt nach Informationen darüber suchen, wie sie ihre Krankheit bekämpfen bzw. ihren Alltag besser gestalten können. Weniger wichtig sind ihnen dagegen detaillierte Informationen über die Krankheit selbst. Darüber hinaus ist es so, dass gerade in einer emotional aufgeladenen Situation, wie es die Belastung durch eine Krankheit ist, die Informationsaufnahme problematisch sein kann und durch die Berücksichtigung der Situation und des

emotionalen Zustands des Nutzers eine verbesserte Rezeption der Inhalte erreicht werden kann [Zini und Ricci, 2012].

Der entscheidende Punkt dabei ist, dass der Wert und die Bedeutung von Informationen immer relativ zu den Erwartungen des Nutzers sind [Wang und Soergel, 1998]. Diese Erwartungen werden aber von Faktoren *außerhalb* des Systems mitgeprägt, also etwa seinem Vorwissen und damit dem Neuheitswert der Inhalte [Wang und Soergel, 1998], aber auch von affektiven Faktoren [Kuhlthau, 1991]. Diese Prägung der Erwartungshaltung spricht für eine Auseinandersetzung mit den konkreten Auswirkungen auf die Erwartungen der Nutzer.

Singh et al. [Singh et al., 2013] betonen darüber hinaus, dass die Komplexität und Heterogenität der Inhalte im World Wide Web *allgemein* so hoch ist, dass die inhaltliche Berücksichtigung des Nutzerkontexts und seiner situativen Interpretation und Nutzung der jeweiligen Inhalte dringend notwendig ist, um die Informationssuche zu verbessern [auch Shen, 2007].

Untersuchungen zu individuellen Einflüssen der Suchsituation auf das Nutzerverhalten im Web (siehe z.B. [Lindley et al., 2012; Wen et al., 2007; White und Kelly, 2006 oder Byström und Järvelin, 1995]) zeigen ebenfalls die allgemeine Relevanz der subjektiven Bewertung der Inhalte durch den Nutzer in Abhängigkeit seiner Situation.

Da die Wikipedia in wichtigen Punkten ähnliche Eigenschaften wie das Web hat (Heterogenität in Medientypen, Inhalten und der Nutzung [Adafre und Rijke, 2006]) werden diese Annahme hier auf die Wikipedia übertragen. Auch die genannte Problematik, dass die Stichwortsuche nur wenige Informationen über die tatsächlichen Bedürfnisse und Erwartungen des Nutzers transportiert [White und Kelly, 2006], liegt hier vor.

Die Suche in der Wikipedia ist, wie generell die Suche in digitalen Informationssystemen, ein Teil unseres Alltags geworden [Lindley et al., 2012]. Dabei ist nicht klar, wie sich die verschiedenen Situationen auf Suchstrategien und Wunsch der Nutzer auswirken. Welche Situationen gibt es? Wie unterscheiden sie sich? Wie beeinflussen die Situationen die Nutzer? Welche Inhalte werden in der Wikipedia für welche Probleme herangezogen? Welche Rolle spielen dabei die Ideen und Vorstellungen der Nutzer?

---

Welchen Einfluss hat die Informationspräsentation auf die Zufriedenheit und auf den Sucherfolg?

Die Mehrdimensionalität des Nutzerkontextes [Kuhlthau, 1991, Wang und Soergel, 1998, Ingwersen und Järvelin, 2005] erfordert eine umfassende Betrachtungsweise der Suchsituation inklusive der affektiven und kognitiven Aspekte, um den genannten Fragestellungen nachzugehen. Deren Beantwortung ist die Grundlage für eine Verbesserung der Informationspräsentation durch die Berücksichtigung situativer Nutzerpräferenzen.

Mit dem Fokus auf der deutschsprachigen Wikipedia wird diesen Fragen deshalb in dieser Arbeit nachgegangen.



## 4 Abgrenzung des Forschungsgebiets

Diese Arbeit konzentriert sich auf die Nutzungsszenarien beim Informations-Suchverhalten, den Nutzen und die Begründung für Adaptivität und Suchstrategien sowie subjektive Relevanzurteile und die situationsabhängige Rezeption von Inhalten bei der Suche in der deutschsprachigen Wikipedia. Im Folgenden wird ein kurzer Überblick über verwandte Themen gegeben, die aber in dieser Arbeit nicht weiter berücksichtigt werden.

### 4.1 Kontext in verschiedenen Forschungsbereichen der Informationswissenschaft

Generell gilt, dass auch in anderen Bereichen, nicht nur im Information Retrieval, eine Interaktion, die auch kontextbezogene Aspekte berücksichtigt, eine Verbesserung der Kommunikation zwischen Mensch und Maschine erwirken kann ([Picard, 1998], [Axelrod, 2004]). In Bereichen wie Computerspiele und Unterhaltungselektronik gibt es diese Ansätze. Auch Musik kann (unter Anderem) der momentanen Gemütsverfassung des Nutzers und der aktuellen Situation entsprechend ausgewählt werden. Wird etwa bei einer Fahrt im Auto nach einem passenden Musikstück gesucht, ist es möglich Kontextfaktoren wie den Verkehr, das Wetter, die in diesem Moment durchfahrene Landschaft und/oder Gemütsverfassung und persönlichen Vorlieben des Nutzers zu berücksichtigen, um damit die Auswahl zu unterstützen und zu optimieren [Baltrunas et al, 2011].

Im *Semantic IR* bzw. auch im Bereich der *Semantic Recommenders* werden Empfehlungen auf der Basis semantischer Relationen gegeben [Lommatzsch et al., 2013]. Bei der Berücksichtigung des semantischen Kontextes liegt der Fokus allerdings auf einer Optimierung der (Ergebnis-) Vorschläge hinsichtlich der *inhaltlichen* Präferenzen der Nutzer, nicht auf einer Anpassung an die aktuellen situativen Gegebenheiten. Im Prinzip handelt es sich dabei um einen Filter, der – je nach persönlicher Gewichtung verschiedener Facetten durch den Nutzer – die Treffermenge erstellt und dabei semantische Beziehungen berücksichtigt.

Auch beim E-Learning oder bei Tutor-Systemen können bessere Ergebnisse erzielt werden, wenn die Maschine zum Beispiel zumindest auf einfache emotionale Zustände reagieren bzw. sich auf den Nutzer einstellen kann ([Picard, 1998; Ramos und de Bra, 2010]).

Der Computer als ständiger Begleiter im Alltag eignet sich dazu, automatisch so viele Kontextinformationen wie möglich über einen Nutzer zu sammeln und sich dadurch möglicherweise ein umfassendes Bild von den Gewohnheiten und Bewegungen des Nutzers zu machen. Nicht nur die meist täglich verwendeten PCs und Laptops, sondern auch kleine und kleinste Systeme können, z.B. verarbeitet in Schuhen, Kleidung oder Schmuck, ohne Unterbrechung Daten über den Nutzer erfassen [Picard, 1998]. Kann man dann aus den über das Verhalten gelieferten Daten gewisse Muster und Konsequenzen ableiten, ist es vorstellbar, ein ganz persönliches, situativ und nutzerbezogen individuell reagierendes Assistenzsystem zu schaffen. Anwendungen und Untersuchungen dazu gibt es etwa im E-Health Bereich, wo kleine, tragbare Computer Informationen über den Nutzer sammeln und auf dieser Basis eine individuelle, nutzerbezogene Beratung zu Ernährung und gesunden Verhaltensweisen (zum Beispiel bei zu wenig Bewegung) geben können [Ludwig et al., 2011]. Systeme, die automatisch Informationen über den Nutzer und seinen Kontext sammeln, z.B. durch kontinuierliches Aufzeichnen von Tätigkeiten, sind nicht Gegenstand dieser Arbeit.

## 4.2 Implementierung affektiver und kontextsensitiver Interfaces

Bei der Integration affektiver und kontextsensitiver Interfaces müssen mehrere Komponenten eines Programmes gut zusammenarbeiten: verschiedene Sensordaten sollen zuverlässige Informationen über das aktuelle Verhalten des Nutzers liefern und etwaige Voreinstellungen zu grundsätzlichen Präferenzen müssen berücksichtigt werden. Aus diesen Informationen werden dann auf Basis der Erkenntnisse aus der Forschung Regeln erstellt und die entsprechenden Schlüsse gezogen, so dass ein System die Inhalte korrekt auswählen und präsentieren kann. Die Inhalte selbst müssen wiederum gekennzeichnet sein, um ein *matching* mit den über den Nutzer erfassten Informationen zu ermöglichen. Die Implementierung einer solchen Benutzerschnittstelle ist nicht Teil



dieser Arbeit. Die Erkenntnisse und Untersuchungen dieser Arbeit aber sind die Grundlage für eine mögliche Realisierung einer kontextsensitiven Informationspräsentation.

### 4.3 Beschränkung auf eine Auswahl an Einflussgrößen

Es zeigt sich in der Literatur und auch im Laufe dieser Arbeit, dass es eine große Anzahl von Faktoren gibt, die die individuellen und subjektiven Bedürfnisse und Wünsche der Nutzer, vor allem bei der Informationssuche im Alltag beeinflussen. Diese Liste ist schwer abgrenzbar, da einzelne Faktoren wie z.B. die Stimmung wiederum dem Einfluss anderer, nicht direkt mit dem Suchtask verbundenen Bedingungen, wie etwa dem Wetter oder der Jahreszeit, unterliegen.

Deshalb ist es wichtig, einen klaren Bezugsrahmen der Untersuchungen festzulegen. Konkret bedeutet das, dass nicht nach dem Ausschlussprinzip (jeder in der Welt vorhandene Faktor wird auf Einfluss getestet) vorgegangen wird, sondern dass Faktoren berücksichtigt werden, bei denen nachvollziehbare und schlüssige Hinweise auf mögliche Einflüsse vorliegen.

Dieses Vorgehen entspricht dem Vorschlag von [Benyon et al., 1994]. Er betont ebenfalls, wie wichtig es sei, nur Faktoren zu verwenden, die

- a) wirklich einen *Einfluss* auf die Nutzerpräferenzen haben und
- b) auch sinnvoll *erfasst* werden können.

### 4.4 Datenerhebung anhand der deutschsprachigen Wikipedia

Der Schwerpunkt der Arbeit ist die Konzentration auf die Informationssuche und Informationsrezeption in der deutschsprachigen Wikipedia, denn die Wikipedia ist aufgrund ihres Bekanntheitsgrades, der Anknüpfung an das traditionelle Genre der Enzyklopädien und ihrer vielfältigen Inhalte und Inhaltselemente ein System, bei dem erstens die Vermutung der verschiedenen Nutzungsszenarien plausibel scheint und zweitens die Annahme besteht, dass eine Berücksichtigung dieses Potentials zur Verbesserung der Informationspräsentation und einer höheren Nutzerzufriedenheit birgt.

Darüber hinaus ist generell eine Anwendung der kontextrelativen Informationspräsentation nur in einem geschlossenen System realistisch vorstellbar, da weitere Voraussetzungen wie entsprechende Anreicherung mit Metadaten und eine mehr oder weniger einheitliche Qualität und Verfügbarkeit von Inhalten dazu notwendig sind.

Bei der Wikipedia ist eine Qualitätssicherung gegeben (siehe dazu Kapitel 9.6). Die Themen in der Wikipedia sind vielfältig, es sind realistische, alltagsnahe Szenarien vorstell- und für die Datenerhebung konstruierbar. Darüber hinaus gibt es in der Wikipedia ein endliches und überschaubares Set an Inhaltstypen, die bereits bewährt sind<sup>4</sup>. Somit können konkrete Analysen der Rezeption dieser Elemente durchgeführt werden. Die Wikipedia ist eine bekannte und viel genutzte Website (siehe dazu auch Kapitel 11.3.1), das erleichtert für die empirischen Datenerhebungen die Bearbeitung der Aufgaben durch die Probanden und deshalb findet die empirische Datenerhebung ausschließlich mit und in der Wikipedia statt.

Eine Anwendung der Forschungsfragen und -ergebnisse auf das Gebiet des adaptiven Hypertextes wäre ebenfalls vorstellbar. Hier ist eine stärker auf die Bewegungsmuster der Nutzer im Hypertext zugeschnittene Fragestellung denkbar. Hypertextsysteme sind ebenfalls Gegenstand von verschiedenen Nutzungsszenarien und es existiert somit auch die Problematik, dass sie unter Umständen den stark unterschiedlichen Anforderungen nicht gerecht werden können [Brusilovsky et al., 2000]. Auch hier bekommt jeder Nutzer zunächst dasselbe Angebot an Inhalten, unabhängig von seinem Interesse, Vorwissen und Ziel. Infolgedessen gibt es auch hier seit den 1990er Jahren Überlegungen zu adaptiven Systemen [Brusilovsky et al., 2000], um diese Problematik zu beheben. Dabei ist hier außer dem Interesse und den Zielen des Nutzers auch die interne Struktur verantwortlich für eine gute Navigation – die Navigationsziele sind weiter abhängig von den bereits besuchten Inhalten und rezipierten Informationen. Ansätze wie eine Personalisierung von Inhalten, Navigationsassistenten, spezielle Anwendungen wie aus dem E-Health oder dem E-Commerce Bereich so wie Lern- und

---

<sup>4</sup> Das ist wichtig, da so der Einfluss durch die grundsätzliche Qualität der Inhaltstypen an sich schwächer ist.

Taggingverfahren zur Optimierung von Hypertextstrukturen sind dargestellt in [Brusilovsky et al., 2000]



## 5 Unterschiedliche Nutzungsszenarien

Die erste wesentliche Frage dieser Arbeit ist die nach verschiedenen Nutzungsszenarien im Information Retrieval im Allgemeinen und bei der Wikipedia im Speziellen.

Zur Klärung dieser Frage wird zunächst, als eine wesentliche Begründung für das Entstehen stark unterschiedlicher Nutzungsszenarien, dargestellt, dass die Informationssuche in digitalen Systemen ein Teil unseres Alltags geworden ist. Es wird generell auf die Forschung zum Informationsverhalten und zu den Komponenten eines Suchprozesses eingegangen, um zu evaluieren, an welchen Stellen des Suchprozesses die Unterschiede im Verhalten und den Anforderungen der Nutzer liegen können. Zur weiteren Erklärung der Entstehung unterschiedlicher Nutzungsszenarien werden Unterscheidungen zwischen verschiedenen Suchauslösern, Suchzielen und Motivation für eine Suche besprochen, sowie zusätzliche mögliche Faktoren, die die subjektive Relevanz und Rezeption von Information der Nutzer sowie deren Suchverhalten beeinflussen.

### 5.1 Suche als Teil unseres Alltags

*„From the early days of computers, search has been a fundamental application that has driven research and development. [...] As computers have become consumer products and the Internet has become a mass medium, searching the Web has become a daily activity for everyone from children to research scientists.“*

[Marchionini, 2006, S.1]

Dieses Zitat betont einen wesentlichen Punkt: Die Informationssuche im Web ist für Jung und Alt ein Teil unseres Alltags geworden.

Die Suche nach Informationen ist demnach auch nicht mehr nur auf eine Recherche zu Arbeitszwecken beschränkt. Sie ist schon seit den neunziger Jahren und noch einmal ganz verstärkt seit Beginn des 21. Jahrhunderts eine alltägliche Beschäftigung geworden [Spink und Cole, 2001]. Längst gibt es nicht nur mehr „die Experten“ in der Infor-

mationsrecherche. Gesucht wird von jedem, zu jeder Zeit und nach jedem vorstellbaren Ziel [Lindley et al., 2012; Marchionini, 2006].

Savolainen [Savolainen, 1995] beschreibt eine Unterscheidung zwischen dem sog. *work-based* und *non-work-based information seeking*<sup>5</sup>. Dabei bezieht sich Savolainen auf die Informationssuche einerseits im Rahmen einer Arbeit oder Erwerbstätigkeit und andererseits dem *ELIS*, dem *everyday life information seeking* [Spink und Cole, 2001]. Diese Form der Informationssuche, die Teil des Alltags und nicht des Arbeitslebens ist, unterscheidet sich von der in einem „kontrollierten“ [Spink und Cole, 2001] Umfeld stattfindenden Informationsrecherche. Damit ist eine weitere Form der Informationssuche [ELIS, Spink und Cole, 2001] entstanden.

In dieser alltäglichen Informationssuche weichen die Nutzer von bewährten und erlernten Suchkonzepten und -strategien ab. Je nach verfügbarer Zeit und Möglichkeiten werden neue, weniger oder nicht mehr kontrollierte Umgebungen in die Informationssuche mit eingebunden [Spink und Cole, 2001]. Damit wird die Informationssuche im Alltag verstärkt Gegenstand weiterer „Umwelteinflüsse“ und bzw. des Einflusses des sozialen Kontexts, in dem sie ausgeführt wird. Sowohl Motivation des Suchenden, die benutzten Inhalte, sein intendiertes Ziel als auch seine Vorstellung der Zielerfüllung gewinnen damit an Einfluss [Spink und Cole, 2001].

Das ist dadurch zu erklären, dass jede soziale Aktion, in diesem Fall eine Informationssuche als Teil des Alltags, in einem Kontext entsteht und stattfindet [Johnson, 2002]. Es besteht keine abstrakte, theoretische Vorstellung und kein prototypisches Vorgehen bei der Problemlösung, sondern die tatsächliche Charakterisierung einer Suchaktion entsteht durch ihren Kontext.

Da die Informationssuche im Web eher autodidaktisch erlernt wird, entwickelt der Nutzer jeweils eigene Strategien. Die klassischen Modelle und die Forschung zur Informationssuche berücksichtigen den Aspekt des *everyday life information seeking* [Spink und Cole, 2001] oft wenig [McKenzie, 2003], vor allem die inhaltlich weniger zielorientierte Suche, das in dieser Arbeit als *browsing* oder *exploratory search* bezeichnete Verhalten (siehe dazu auch 5.5), das auch schon Wilson [Wilson, 1977] in Bezug auf Zeitun-

---

<sup>5</sup> Hier wird derselbe Term verwendet wie später bei der Unterscheidung nach den verschiedenen Motivationen für eine Suche, allerdings ist die Bedeutung nicht dieselbe. In diesem Fall meinten die ursprünglichen Autoren [Savolainen, 1995 und Spink und Cole, 2001] tatsächlich auf die Arbeit im Sinne von Erwerbstätigkeit bezogene Informationsrecherchen, wogegen später derselbe Term verwendet wird, um zielloses Browsen von einem tatsächlichen Informationsbedarf abzugrenzen.

gen oder Fernsehen beschrieb (Blättern in Magazinen, „Zappen“ durch die Fernsehprogramme). In Kapitel 5.4 wird auf diese spezielle Fragestellung detailliert eingegangen.

## 5.2 Information-Seeking und Information Behaviour

Im Forschungsbereich der Informationssuche und des Information Retrieval existieren Modelle, die mit verschiedenen Schwerpunkten den IS&R<sup>6</sup> Prozess, die Interaktionen von Nutzer und System sowie deren Auswirkungen darstellen. Eine gute Übersicht dazu findet man bei Ingwersen und Järvelin [Ingwersen und Järvelin, 2005, S. 195 - 256]. Diese Modelle konzentrieren sich sowohl auf den Nutzer und seinen Suchprozess als auch auf die systemseitige Sichtweise, die Abbildung und den Abgleich von Inhalten.

Das Vorgehen bei der Suche nach Informationen ist dabei ein Prozess, den jeder je nach Fähigkeiten und Erfahrungen individuell gestaltet. Er lässt sich aber für ein Grundverständnis abstrahiert und verallgemeinert darstellen. Es gibt in der Forschung theoretische Modelle, z.B. von [Marchionini, 1989; Kuhlthau, 2004; Belkin, 1980 oder Ingwersen und Järvelin, 2005], die dies gut veranschaulichen. Um den Suchprozess zu verstehen und die situationsrelativen Elemente einer Suche zu identifizieren, ist es notwendig, diese strukturellen Erkenntnisse zu berücksichtigen.

Marchionini hat bereits 1989 die Fragen gestellt:

*„3. What are the relationships among user, task and search pattern?*

*4. What search patterns are exhibited and how are these patterns related to information-seeking strategies?“*

[Marchionini, 1989, S.3]

Er hält fest, dass die Informationssuche an sich eine ganz bestimmte Instanz des allgemeinen Handelns des Problemlösens ist [Marchionini, 1989]. Der Nutzer hat ein Ziel

---

<sup>6</sup>Information Seeking and Retrieval

(=task), einen oft eher implizit vorhandenen Plan, dieses zu erreichen und er korrigiert notfalls während der Ausführung des Plans immer wieder seine Richtung, wenn er merkt, dass er eventuell im Moment nicht zielführend arbeitet (=search pattern). Dabei wird der Nutzer immer geleitet von

1. seinem gewünschten Ziel
  2. seinen kognitiven und affektiven Fähigkeiten und seinem Vorwissen und
  3. davon, was das jeweilige System ihm in diesem Moment zu bieten hat
- [Marchionini, 1989].

Deshalb ist es notwendig, die konkreten Faktoren zu untersuchen, die die Entscheidungen der Nutzer beeinflussen, um ihre Strategien zu verstehen.

*Information Behaviour* und *InformationSeeking Behaviour* sind als eine Abfolge von Entscheidungen zu betrachten, die durch die Faktoren aus dem Kontext des Nutzers, inklusive seiner kognitiven Struktur und dem aktuellen Informationsbedürfnis und den Reaktionen auf die bereits evaluierten Inhalte, beeinflusst werden (so auch [Wilson, 1999]).

In [Fischer et al., 2005] werden ausführlich über 70 konzeptionelle Frameworks zur Informationssuche und der Frage, wie sucht, teilt und verwaltet der Mensch seine Information, beschrieben. Die dort besprochenen Modelle beziehen sich auf unterschiedliche Ebenen und Aspekte der Suchszenarien, dabei auch ethnographische oder speziellere Bereiche wie *Small-World Network Exploration*. Im Zusammenhang mit dieser Arbeit sind vor allem die zentralen Konzepte eines *Suchablaufs* inklusive *Ziel* und *Motivation* relevant. Der Einfluss dieser Faktoren auf die Präferenzen der Nutzer bei der Suche in der Wikipedia ist – im Sinne der Zusammenhänge *among user, task and search pattern* [Marchionini, 1989, S.3] – noch ungeklärt.

Das folgende Modell von Hearst [Hearst, 2009] zeigt das Grundprinzip des Suchprozesses (siehe Abbildung 5-1). Was hier als *Task* beschrieben wird, ist in dieser Arbeit als Entstehungssituation für einen Suchauslöser definiert. Aus dem jeweiligen *Task* ergibt sich ein *Info Need*. Dieses wird in dieser Arbeit konkret durch das inhaltliche Suchziel,



und zusätzlich durch bereits vorhandenes Vorwissen und weitere Präferenzen des Nutzers determiniert. Die in dem Modell von Hearst aufgeführte *verbal form* entspricht hier der Übersetzung des Informationsbedarfs in eine Stichwortsuche bei der Wikipedia, die die *query* darstellt. Der vorhandene Suchalgorithmus führt einen Abgleich mit den vorhandenen Dokumenten durch und liefert die entsprechende Auswahl an den Nutzer zurück. Je nach Zufriedenheit mit den Ergebnissen verändert der Nutzer dann seine Suchanfrage und wiederholt den Prozess ab der (neuen) *verbal form* seines Informationsbedarfs.

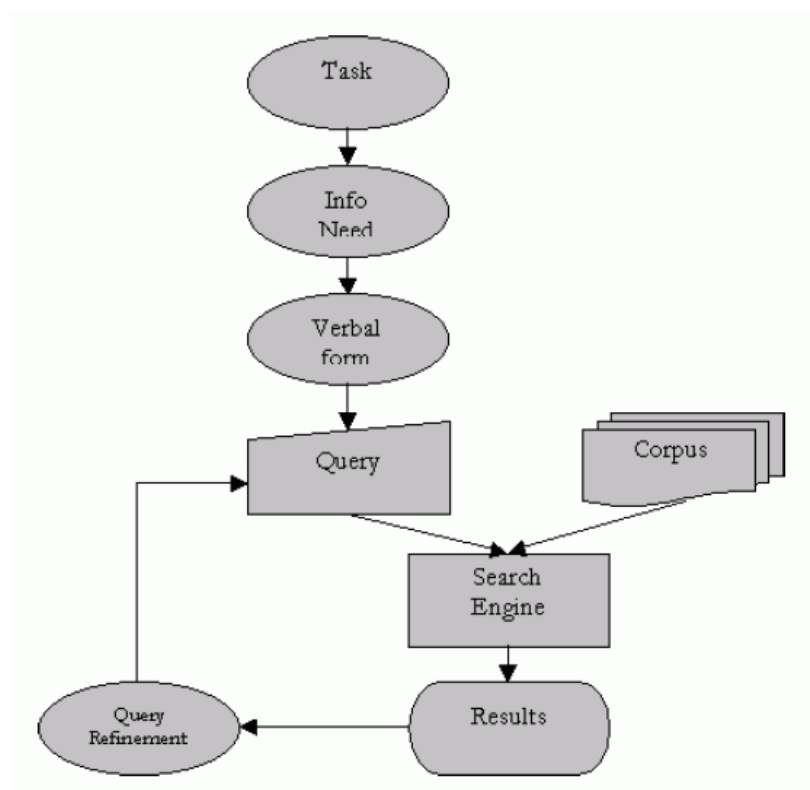


Abbildung 5-1 Standardmodell des Suchprozesses von [Hearst, 2009, Kapitel drei]

Bei der Suche in der Wikipedia übermittelt der Nutzer sein Suchziel in der Regel in Form einer *query* (sollte er nicht eine andere Einstiegsform wählen). Er müsste – sollten seine persönlichen Vorstellungen berücksichtigt werden – damit theoretisch sein Ziel und seine Vorstellung davon, wie er dieses erreichen will, in Form der gewählten Terme an das System übermitteln. Das ist aber nur sehr begrenzt oder nicht möglich [White und Kelly, 2006]. Der Nutzer muss genau diejenigen Informationen finden, die in sein mentales Modell der Welt und speziell zur Problematik, die er durch Ausführen

der Suche zu lösen versucht, passen. Er wird die Ergebnisse seiner Suchanfrage dahingehend evaluieren und bewerten [Wang und Soergel, 1998]. Ziel dieser Arbeit ist es, diese Bewertung durch empirische Methoden explizit zu machen, denn dieser Bestandteil des hier beschriebenen *Information Seeking Behaviour* kann durch die Interaktion des Nutzers beobachtet werden: Welche der angebotenen Elemente nutzt er für seine Problemlösung?

Die einzelnen Schritte des allgemeinen Modells des Suchprozesses bleiben dabei gleich, da sie nur das Handeln selbst betreffen und abstrakt beschrieben sind. Die konkreten Inhalte (vom *Information Need* bis hin zur Komplettierung der Suche) sind dagegen variabel.

Manche Modelle (wie z.B. das *Berry Picking Modell* von [Bates, 1989]) gehen von einem konstanten Verhalten und vor allem von einem konstanten Informationsbedürfnis über den kompletten Suchzeitraum aus. Dies würde der formalen Einteilungen der *Information Needs* nach z.B. [Marchionini, 2006] folgen. Andererseits liegt die Vermutung nahe, dass sich während eines Suchprozesses durch die Interaktion mit dem System und den Ergebnissen selbst sowohl Suchstrategie als auch das Suchziel ändern (z.B. durch Konkretisierung, Ablenkung, Auffinden von Informationen, nach denen nicht gesucht wurde, da deren Existenz nicht vermutet wurde etc.). Besonders bei der in 5.1 und 5.5 beschriebenen Freizeitsuche kann – wie von McKenzie [McKenzie, 2003] beschrieben – das *info need* auch weniger systematisch und definiert sein. Es ist also festzuhalten, dass beide Möglichkeiten eintreten können, sowohl ein über die Suche konstantes Suchziel als auch Wechsel während des Prozesses.

Zu dieser Frage beschreibt Carol Kuhlthau mit Hilfe ihres ISP (*Information Search Process*) Modells [Kuhlthau, 1991] nachvollziehbar verschiedene Zustände eines Suchprozesses. Der Nutzer durchläuft demnach sechs Stufen, die Kuhlthau als

1. Feststellen eines Informationsbedarfs (vergleichbar dem ASK von [Belkin, 1980]),
2. einer genaueren Identifizierung des Topics,
3. der Suche nach der gewünschten Information im identifizierten Bereich,

4. einer stärkeren Fokussierung<sup>7</sup>,
5. einer Auswahl der gefundenen Informationen und dem Zusammentragen der zur Erfüllung des Informationsbedürfnisses notwendigen Teile davon und
6. dem Abschluss der Suche durch Erfüllung des Informationsbedürfnisses

definiert. Diese Einteilung entspricht der Hypothese, dass sich durch die Interaktion mit den Inhalten der kognitive Zustand des Nutzers ändert und dadurch auch sein Zustand im Suchprozesse, so auch [Ingwersen und Järvelin, 2005].

Wilson [Wilson, 1999] stellt unter dem Aspekt des *information behaviour* noch weitere Bereiche um das reine *information-searching behaviour* (Abbildung 5-2) dar. Dabei berücksichtigt er ebenfalls die Frage, ob der Nutzer mit den jeweils gefundenen Inhalten zufrieden ist und damit seine Suche erfolgreich war. Auch dieses Modell verdeutlicht den Untersuchungsgegenstand dieser Arbeit: die hier allgemein bezeichneten Bereiche wie „Need“ oder „Satisfaction or Non-satisfaction“ beschreiben abstrakt die Nutzungsszenarien und die davon abhängige situative Zufriedenheit der Nutzer. Diese wiederum beeinflussen seine Suchstrategien [Wilson, 1999].

---

<sup>7</sup> Vergleichbar einer immer genaueren Vorstellung der gewünschten Information, die am Anfang nur in Ausnahmefällen, z.B. unter Umständen beim sog. *fact finding* gegeben ist, aber auch dann nicht unbedingt, denn in der Regel ist das Informationsbedürfnis nicht zwingend explizit vorhanden.

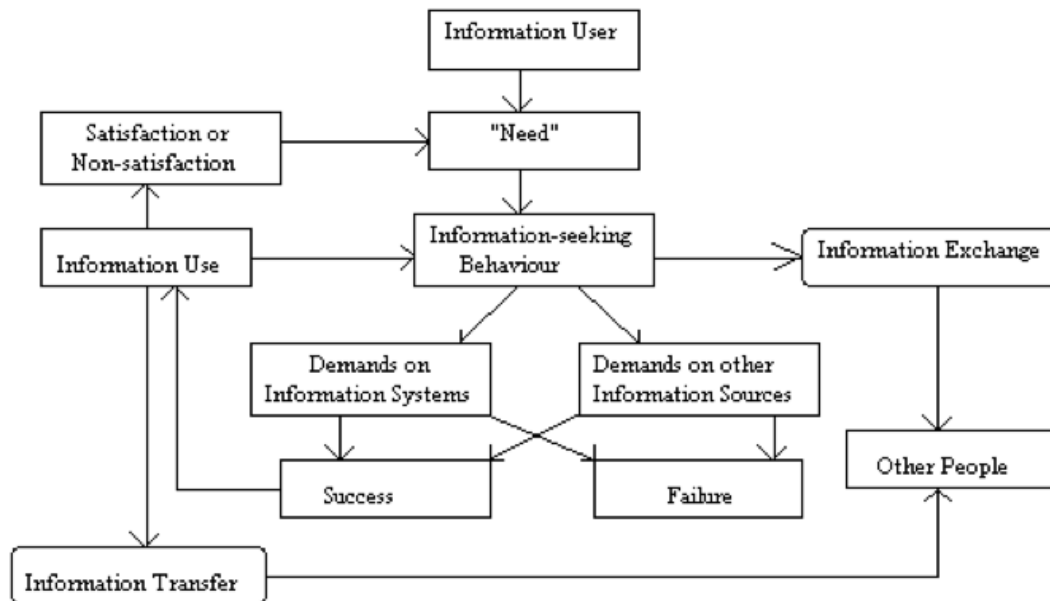


Abbildung 5-2 Information Behaviour Model von [Wilson, 1999, Kapitel 2]

Die in der Forschung vorhandenen Modelle entspringen entweder einer theoretisch-analytischen Vorgehensweise oder basieren auf empirischen Daten und Erfahrungen. Im Laufe der Zeit wurden Faktoren identifiziert, die einen Einfluss auf das Verhalten und die Wahrnehmung der Nutzer im Rahmen der Informationssuche haben (vgl. dazu verschiedene Modelle, nachzulesen in [Ingwersen und Järvelin, 2005]).

Ingwersen und Järvelin ergänzen dabei diese Modelle vor allem um den *kognitiven* Aspekt, der den Nutzer beeinflusst, also seine Wissensstruktur und deren Einfluss auf die Rezeption von Inhalten. Ebenso fokussieren sie die Thematik unterschiedlicher Arten von Suchanfragen und deren Komplexität.

Das nachfolgende Modell aus [Ingwersen und Järvelin, 2005] zeigt, dass der Kontext im *world model* des Nutzers eine Rolle spielt. Dieser wird demnach beeinflusst von Faktoren, die das System nicht kennt. Der Generator der Nachricht hat ebenfalls einen Kontext<sup>8</sup>, dieser geht aber beim Transport der Nachricht verloren. Der Empfänger interpretiert die Nachricht bzw. die Information in seinem *eigenen* Kontext (bestehend aus seiner vorhandenen kognitiven Struktur, seiner Problemlösungsvorstellung etc., da

<sup>8</sup> In diesem Fall das System – Kontext ist dann vermutlich rein semantisch zu sehen.

der Nutzer wiederum den Kontext des Generators nicht kennt). Der Informationsbegriff im Bezug zum *cognitive viewpoint* [Ingwersen und Järvelin, 2005] wird dabei auf Basis der Definition von Information von ([Steinmueller, 1993] u.a., ähnlich auch bei [Ingwersen, 1992; Wersig, 1971]) verwendet: Information ist zweckrelativ und subjektrelativ.

Im Falle einer Suche in der Wikipedia ist der Kontext des Generators der Abgleich der Suchterme, der Kontext des Nutzers sein aktuelles Nutzungsszenario. Ingwersen und Järvelin [Ingwersen und Järvelin, 2005] behandeln hier allerdings den Begriff Kontext bewusst als offenes Konzept, als generelle Faktoren aus der Umwelt, die den Rezipienten beeinflussen. Diese Faktoren werden im empirischen Teil dieser Arbeit und dem hier erstellten Modell wiederum aufgegriffen und operationalisiert.

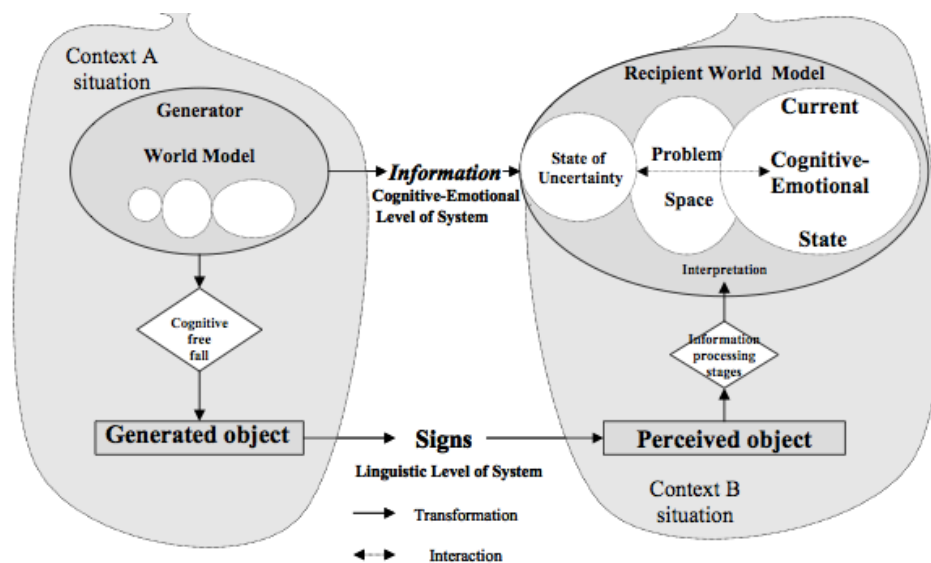


Abbildung 5-3 The cognitive communication system for Information Science, information seeking and IR. Revision of Ingwersen [1992, p.33; 1996, p.6], from [Belkin, 1978], [Ingwersen und Järvelin, 2005, S.39]

Ein weiteres Modell aus [Ingwersen und Järvelin, 2005, S.277] beschreibt einen *generellen* Aufbau von Nutzer, Kontext und System (siehe Abbildung 5-4). Es zeigt, dass der Nutzer umgeben von seinem Kontext ist und von dieser „Einbettung“ aus mit dem Interface interagiert. Seine Interaktion wird vom Kontext beeinflusst und somit findet ebenfalls ein – indirekter – Einfluss auf die Informationsobjekte durch die Beurteilung und Bewertung der Objekte statt. Darüber hinaus sind auch die potentiellen Informa-

tionsobjekte und die Interaktion mit diesen wiederum selbst Einflussfaktoren, die den Kontext (hier vor allem im Sinne der kognitiven Struktur des Nutzers, aber auch auf affektiver Ebene) selbst wieder beeinflussen [Ingwersen und Järvelin, 2005]. Dieses Modell zeigt die komplexen Zusammenhänge zwischen Informationsobjekten, Nutzer und der Situation, die vielfach bidirektional bestehen. Konkret bedeutet das, dass Ursache und Wirkung nicht immer klar getrennt werden können; so hat der Nutzer zwar einen bestimmten Zustand, wenn der Task beginnt, aber das Suchverhalten kann diesen Zustand selbst insofern beeinflussen, als dass dieser sich durch den Verlauf der Suche ändert (worauf sich das Suchverhalten wiederum dem neuen Zustand anpasst). Dies ist auch entsprechend [Kuhlthau, 1991]. Auch das Modell Abbildung 5-4 beschreibt abstrakt den Bezugsrahmen der Untersuchungen dieser Arbeit und bildet eine Basis für konkrete Überlegungen zu Suchverhalten und Interaktion mit der Wikipedia.

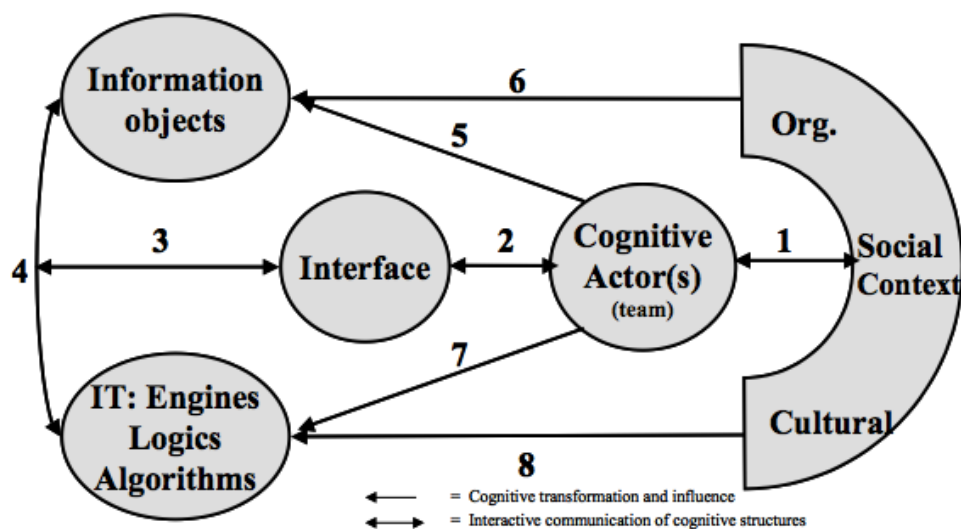


Abbildung 5-4 Interactive Information Seeking, Retrieval and Behavioral processes [Ingwersen und Järvelin, 2005, S.277]

Bereits 1995 haben Byström und Järvelin [Byström und Järvelin, 1995] ein Modell entworfen, das ebenfalls den Aspekt des Kontexts und der Nutzeraktionen in Bezug setzt, ohne allerdings die einzelnen Faktoren jeweils „auszuformulieren“ und konkrete Erfassungsbereiche oder Aktionen festzulegen. Allerdings zeigt das Modell gut, dass das Problem der Wechselwirkung zwischen Kontext und Nutzer mehrdimensional ist, da nicht nur auf einer Ebene sondern auf vielen Ebenen Einflussfaktoren und Reaktionen

darauf vorhanden sind. Situative und persönliche Faktoren spielen demnach [Byström und Järvelin, 1995] eine Rolle für die konkrete Analyse des Informationsbedarfs, der wiederum ausschlaggebend ist für Relevanz und Ranking der Ergebnisse eines IR-Vorgangs.

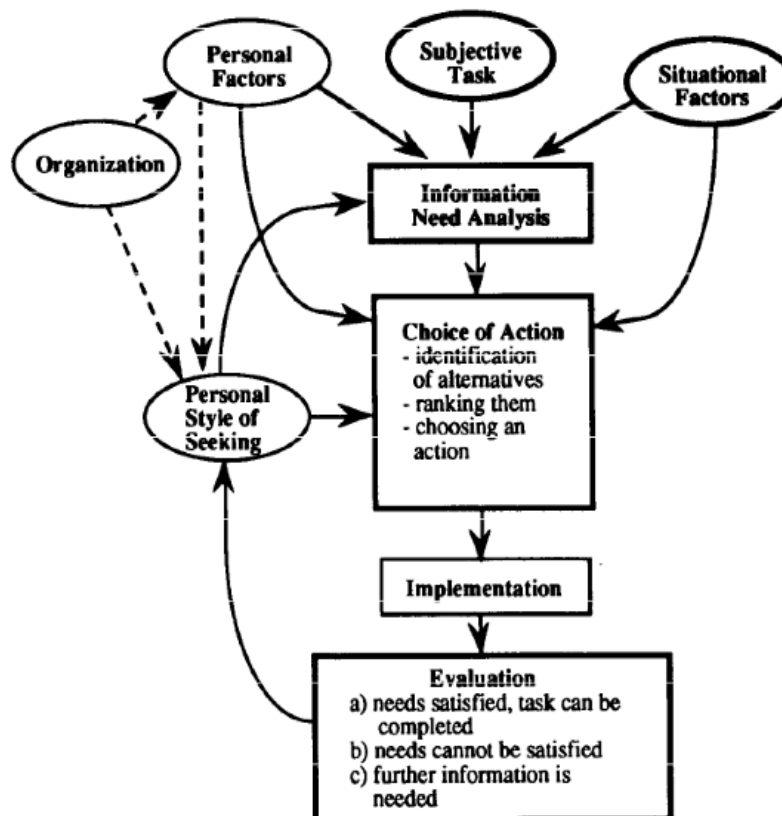


Abbildung 5-5 Information Seeking Model von Byström und Järvelin [Byström und Järvelin, 1995, S.7]

Fasst man die Theorien und Modelle zum *Information Seeking* Prozess zusammen, so ergeben sich folgende wesentliche Elemente des Suchprozesses (nach [Marchioni, 1989; Hearst, 2009]):

Der Nutzer erkennt zunächst, dass er ein „Problem“ (i.d.R. in Form eines Informationsdefizits (*anomalous state of knowledge* [Belkin, 1980] oder in sog. *casual leisure* Szenarien auch anderer Natur wie z.B. Langeweile) hat. Im nächsten Schritt muss das „Problem“ konkretisiert und für eine Suchanfrage an ein Informationssystem verbalisiert werden. Dabei ist entscheidend, dass unter Umständen Teilaspekte des Suchproblems

durch die Verbalisierung nicht erfasst werden (z.B. Umfang, Modus oder Detailschärfe der erwarteten Ergebnisse). Im nächsten Schritt des Suchprozesses evaluiert der Suchende die vom System auf seine Query gelieferten Ergebnisse entsprechend seinen persönlichen Vorstellungen und wiederholt die letzten zwei Schritte bei Bedarf [Hearst, 2009]. Wenn der Nutzer entsprechend seinem initiativen Bedürfnis ausreichend Information aggregiert hat beendet er die Suche (hierzu auch [Kulthau, 1991]). Das Zusammenspiel folgender Konzepte aus dem hier dargestellten Suchprozess wird in dieser Arbeit als Nutzungsszenario definiert:

- Task (Kapitel 5.3)
- Information Need (Kapitel 5.4)
- Information Use (Kapitel 5.5, Kapitel 7.2.3, ab Kapitel 11.2)
- Evaluierung der Ergebnisse/Bewertung der Ergebnisse (Kapitel ab 11.3)
- Zufriedenheit oder Unzufriedenheit (ab Kapitel 7.2, 12.3.5)
- Interaktion mit den Inhalten (ab Kapitel 11.3)

Es wird anschließend der Frage nachgegangen, in welchen unterschiedlichen Ausprägungen diese Konzepte vorkommen und wie sich diese wiederum auf die Präferenzen der Nutzer bei der Informationssuche in der Wikipedia auswirken.

### 5.3 *Searching for a reason* – Warum sucht der Mensch?

Wilson [Wilson, 2006] kritisiert, dass eine globale Verwendung des Terms *information need* den vielen unterschiedlichen Einzelfällen und Ausprägungen nicht gerecht wird. Es ist sinnvoll, zwischen den *human needs* [Wilson, 2006, S. 6] und dem konkreten *information need*, der daraus entsteht, zu unterscheiden (Bsp: Person hat Hunger → *human need*, Person sucht nach einem Kochrezept → *information need* ).

Deshalb steht zunächst die Frage im Vordergrund, warum sich der Mensch auf die Suche nach etwas macht. Neben dem inhaltlichen Informationsziel ist die Situation, die den Bedarf nach bestimmten Inhalten auslöst, hier also relevant. Dieser Teil der Fragestellung bezieht sich vor allem auf den oben beschriebenen *task* nach [Hearst, 2009], der initiativer Bestandteil des Suchprozesses ist.



Marchionini [Marchionini, 1989, S.2] beschreibt zunächst allgemein „*Search*“ als eine „*fundamental life activity*“.

Etwas, oder nach etwas zu suchen ist genuin menschlich, sozusagen ein Grundbedürfnis. Woher kommt dieses Grundbedürfnis?

Nach der Maslowschen Bedürfnis-Hierarchie [Maslow, 1943] versuchen alle Organismen zunächst und an erster Stelle, ihre *physischen* Grundbedürfnisse zu erfüllen. Erst wenn das geschehen ist geht die Suche weiter nach sozialen und psychischen Elementen, die wir benötigen, um unser Leben und unsere Welt zu verstehen. Das Bedürfnis nach Informationen fällt – je nach Auslöser und Ziel – in beide Kategorien. Suchstrategien und -systeme ändern sich zwar, aber die Suche und das Bedürfnis, sich zu informieren, bleiben. Demnach ist der Ursprung einer Informationssuche außerhalb des Informationssystems zu sehen. Eine bestimmte Situation, der *task* [Hearst, 2009] löst das Bedürfnis oder den Bedarf nach Informationen aus. Diese werden dann gesucht.

In einem Projekt unter der Leitung von Barbara Wildemuth, University of North Carolina at Chapel Hill, Luanne Freund, University of British Columbia und Elaine Toms, The University of Sheffield wurde ein *Repository of assigned search tasks* [Wildemuth et al., 2010] erstellt, in dem aktuell (September 2013) 1019 *task types* aufgelistet sind, die jeweils der Forschungsliteratur entnommen wurden. Dort ist eine umfangreiche Menge möglicher suchauslösender Szenarien gesammelt. Daraus wiederum lässt sich die Existenz vieler unterschiedlicher *tasks* nach [Hearst, 2009] ableiten.

Bei den beschriebenen Tasks handelt es sich vor allem um Einzeltypen. Zu deren Definition werden nicht immer die selben Kategorien an Eigenschaften. So ist ein Task Typ etwa eine „*web based search*“<sup>9</sup> – hier fehlen aber Angaben zu Ziel, Umfang und Motivation. „*Well-defined topical information need*“<sup>10</sup> dagegen entspricht eher einer allgemeineren Beschreibung, wogegen „*Where to stay*“<sup>11</sup> als „Tasktyp“ wiederum eine semantische Komponente (die Suche nach Unterkunft) erfasst. Es werden weiterhin *image search*, *news reading task*, *background search*, *know item search*, *work tasks*, *exploratory web search* [Wildemuth et al., 2010] und viele mehr unterschieden. Manche dieser Definitionen

---

<sup>9</sup> Siehe: Systematic Review of assigned search tasks [[http://ils.unc.edu/searchtasks/search\\_all\\_task.php](http://ils.unc.edu/searchtasks/search_all_task.php), abgerufen 18. September 2013].

<sup>10</sup> Ebda.

<sup>11</sup> Ebda.

berücksichtigen bereits Informationen über die Situation selbst hinaus, z.B. auch über das konkrete Suchziel. Dahinter stecken aber unterschiedlichen Situationen, in denen ein Suchauslöser entsteht (z.B. Nutzer benötigt eine Unterkunft, möchte etwas wiederfinden etc.). Das Informationsbedürfnis und die auslösende Situation lassen sich nicht trennen. Um Nutzungsszenarien vollständig beschreiben zu können ist auch die Entstehungssituation notwendig.

Lindley et al. [Lindley et al., 2012] gingen dazu konkret der Frage nach, wie Alltagssituationen aussehen, in denen der Nutzer eine Suche startet. Sie beobachteten Probanden dazu über einen Zeitraum von vier Tagen mittels einer *Diary Study*. Die Teilnehmer hielten jeweils fest, warum sie online gingen und was sie dabei taten. Die Nutzer beschrieben z.B. Situationen wie:

*„if I’m up in the morning the kids have always got their programs on so I sit there in the morning with my phone eating my breakfast and just check the news“*

*„I think when I am coming back [from lunch] I’m still feeling like I want a bit of away time before I really settle in, so part of it’s getting settled back at my computer“*

[Lindley et al., 2012, S. 5]

Was hier deutlich wird, ist die Tatsache, dass die Motivation, um online zu gehen und zu suchen nicht zwingend mit einem konkreten Informationsbedarf verbunden ist. Die Situationen, in denen der Nutzer sucht bzw. eine Informationsrezeption beginnt, sind durch entsprechende persönliche Umstände definiert und werden durch weitere Faktoren charakterisiert. Betrachtet man die hier von Lindley et al. beschriebenen Beispiele, kann man dort entsprechende Informationen isolieren: Uhrzeit [*in the morning*], verwendetes Gerät [*my phone*], weitere Tätigkeiten nebenher [*eating breakfast*] und den Grund für die Suche [*just check the news*]. Der Beschreibung der jeweiligen Situation können so Informationen über den Task entnommen werden. Diese Situation löst nach [Wilson, 1999; Hearst, 2009; Marchionini und Shneiderman, 1988; Marchionini, 2006; Belkin, 1980] dann ein Informationsbedürfnis oder einen Informationsbedarf aus, das oder den der Nutzer zu erfüllen sucht. Eine weitere Unterscheidung von Suchaufgaben

wird hier deshalb nach der Motivation und dem daraus entstehenden (inhaltlichen) Suchziel vorgenommen.

## 5.4 Unterscheidung nach dem Suchziel

Zur Unterscheidung der Suchziele im Information Retrieval gibt es in der Literatur verschiedene Ansätze. Ziel für diese Arbeit ist es, klar abgrenzbare inhaltliche Suchziele zu beschreiben.

Kang und Kim [Kang und Kim, 2003] und Broder [Broder, 2002] unterscheiden die inhaltlichen Ziele einer Suche beide ähnlich nach *informationsorientierten*, *navigationsorientierten* und *transaktionsorientierten* Tasks. Es steht also entweder das Ziel, eine bestimmte Information zu finden und/oder sich über ein bestimmtes Thema zu informieren (informationsorientiert), das Ziel, eine bestimmte Adresse zu finden (navigationsorientiert) oder der Wunsch, eine webbasierte Transaktion durchzuführen (transaktionsorientiert) im Vordergrund. [Kang und Kim, 2003] zeigen auch, dass tatsächlich je nach Bedürfnis, unterschiedliche Suchstrategien verwendet werden. Es ist allerdings zu bemerken, dass erstens kaum eine Suche in nur *eine* der Kategorien fällt und zweitens innerhalb der Kategorien, vor allem aber bei *informationsorientierten* Anfragen, weiter unterschieden werden kann. Da bei der Suche in der Wikipedia vor allem die informationsorientierten Tasks entscheidend sind, ist der Ansatz von [Broder, 2002] hier nicht ausreichend. Gerade die informationsorientierten Tasks müssen weiter unterteilt werden, um den Einfluss des gesuchten Inhalts auf das Nutzerverhalten untersuchen zu können.

Dass innerhalb informationsorientierter Tasks weitere Unterschiede vorhanden sind, zeigen Rose und Levinson [Rose und Levinson, 2004]. Sie differenzieren bei den *informational tasks* weiter nach der Frage, ob eine Suche *open* oder *closed* ist – also dem zentralen Punkt, ob das gesuchte inhaltliche Ziel wohldefiniert oder offen ist. Das wird in dieser Arbeit übernommen. Eine zweite wichtige Unterscheidung, die Rose und Levinson für die informationsorientierten Tasks einführen, ist, ob eine Suche *directed* oder *undirected* ist. Dabei ist der Unterschied vor allem darin zu sehen, ob ein bestimmtes Ziel im Sinne von „Ich möchte genau diese Fakten“ erfahren oder ob ein generelles Ziel

„Ich möchte mich etwas/umfangreicher/ausgiebiger über ein Thema informieren“ vorhanden ist. Auch dieser Aspekt zur Beschreibung eines inhaltlichen Suchziels wird in dieser Arbeit berücksichtigt.

Lee et al. [Lee et al., 2005] nahmen eine automatische Zuordnung von Suchanfragen in die Klassifikation von [Broder, 2002] vor, ließen aber darüber hinaus Testpersonen diese Suchanfragen intellektuell in zwei der drei Kategorien (*informationsorientiert* und *navigationsorientiert*) einordnen. Dabei zeigte sich zwar, dass ein großer Teil der Suchanfragen tatsächlich „vorhersehbare“ Bedürfnisse befriedigen sollte (also auch fremde Personen anhand der *query* richtig einschätzen konnten, welche Ziele die betreffende Person im Sinne der erwähnten Kategorien hatte), *aber* es gab auch Anfragen, bei denen eine eindeutige Zuordnung nicht möglich war. Das bestätigt die Vermutung, dass es Bedürfnisse der Suchenden gibt, die nicht ausschließlich durch die hier verwendeten Klassen beschrieben werden können und es weitere Einflüsse und Charakteristika gibt, die ein Suchziel determinieren.

Das stellten auch Jansen und McNeese [Jansen und McNeese, 2005] fest, die ebenfalls die Klassen von [Broder, 2002] verwendeten. Ihnen gelang automatisch „nur“ eine korrekte Einordnung von 74% der Suchanfragen. Deshalb entschieden sie ebenfalls, weitere Attribute einzuführen, um die Anfragen zu klassifizieren. Für diese Arbeit sind vor allem die ergänzenden Faktoren der Klasse *informationsorientiert* interessant: es zeigte sich, dass auch hier wie bei [Rose und Levinson, 2004] die Berücksichtigung weiterer Aspekte, wie die Unterscheidung zwischen gezielter und ungezielter („*directed*“ vs. „*undirected*“) Suche, die Ergebnisse bei der Zuordnung verbesserten. Das bestätigt, dass eine zusätzliche Unterscheidung der Informationsziele gerade hinsichtlich ihrer Definiertheit und ihrer Endlichkeit die Genauigkeit der Darstellung der Suchziele verbessert.

Weitere Aspekte fügen Calderon-Benavides et al. [Calderon-Benavides et al., 2010] hinzu. Das Ziel dieser Autoren war es, über die „klassische“ Einteilung nach [Broder, 2002] oder auch [Rose und Levinson, 2004] hinaus Faktoren zu erfassen, die das Ziel des Suchenden beschreiben. Es werden hier zusätzlich stärker „inhaltliche“ Faktoren

der Suchanfragen berücksichtigt. So klassifizieren die Autoren z.B. nach Aspekten wie der Genauigkeit, des Themas oder der Aufgabe der Suche. Ebenso berücksichtigen sie, ob es ein zeitliches Limit gibt.

Die Autoren zeigen, dass mit Hilfe der Zuordnung der Anfragen in die bestimmten Bereiche und einer entsprechenden Anpassung des *matchings* die Suchergebnisse tatsächlich verbessert werden konnten. Dies war vor allem der Fall, wenn in Kombination mehrere Klassen zu einer Suchanfrage ausgefüllt wurden, also möglichst viele Informationen über Inhalt und Ziel der Suche vorlagen. Zwischen den Klassen – wie etwa dem Thema und der bevorzugten Quelle – konnten eindeutige Korrelationen nachgewiesen werden [Calderon-Benavides et al., 2010].

Die Untersuchungen von Calderon-Benavides et al. [Calderon-Benavides et al., 2010] zeigen, dass das Ziel einer Suche zusätzlich durch die Spezifizierung des Themas, die Frage, ob es einen abgegrenzten *Scope* [Calderon-Benavides et al., 2010] gibt, sowie durch zeitlichen Restriktionen determiniert wird.

Marchionini [Marchionini, 2006] beschreibt weiterhin den Ansatz, die Suchziele in drei Klassen zu unterteilen. Dabei unterscheiden sich die Klassen, verwendet man die Überlegungen von [Calderon-Benavides et al., 2010] und [Rose und Levinson, 2004], vor allem hinsichtlich des Grads der Definiertheit des inhaltlichen Suchziels und danach, wie endlich die gesuchte Information ist.

Beim Tasktyp *look up* [Marchionini, 2006] handelt es sich um *Fact Retrieval*, beim Typ *learn* um *Knowledge acquisition* [Marchionini, 2006] und bei der dritten Klasse, *investigate* um *Evaluation* [Marchionini, 2006, jeweils Beispiele]. Diese Dreiteilung ist prototypisch zu sehen. In Kombination mit den Erkenntnissen vor allem von [Wildemuth et al., 2010; Broder, 2002; Rose und Levinson, 2004; und Calderon-Benavides et al., 2010] ergeben sich zusammenfassend folgende Variablen, die in dieser Arbeit das inhaltliche Suchziel eines *informationsorientierten* Task beschreiben: **Umfang** und **Endlichkeit** der gesuchten Inhalte und **Grad der Definiertheit des Ziels** bzw. der Genauigkeit der gesuchten Information sowie **Thema** und **zeitliche Restriktionen**. Diese Variablen können in verschiedenen Ausprägungskombinationen vorkommen.

Ein beispielhafter Ansatz zu diesem Thema, der sich auf einen konkreten Kontext bezieht, stammt von Sohn et al. [Sohn et al., 2008]. Um die Herausforderungen der Informationssuche mit mobilen Endgeräten besser bewältigen zu können, führten sie eine Untersuchung durch, um speziell die Ziele bei der Suche unterwegs besser analysieren und verstehen zu können. In einer *Diary Study* befragten sie zwanzig Personen zwei Wochen lang zu ihrem Informationsverhalten mit mobilen Endgeräten. Neben dem konkreten Informationsbedarf („*What was your information need?*“, [Sohn et al., 2008, S.2]) protokollierten die Personen dabei noch „*Where were you?*“ and „*What were you doing?*“ [Sohn et al., 2008, S.2]). Neben dem Informationsbedarf war hier ebenfalls wieder die Situation, in der dieser entstand, wichtig. Die Informationsbedarfe wurden anschließend in 16 Kategorien klassifiziert. Bei diesem Ansatz ist zu erkennen, dass es sinnvoll ist, den jeweiligen Informationsbedarf abhängig vom Kontext zu klassifizieren. Bei der Klassifikation kann ebenso inhaltlich vorgegangen werden.

Eine Klassifizierung von Suchzielen in der Wikipedia entsprechend der aus der Literatur entnommenen Variablen existiert noch nicht. Da aber die Wikipedia, wie bereits beschrieben, ähnliche Eigenschaften hat wie das World Wide Web, auf das sich die genannten Studien beziehen, werden die hier identifizierten Variablen der *information-sorientierten* Tasks auf die Suche in der Wikipedia übertragen.

Eine stärker inhaltlich orientierte Klassifikation von konkreten Informationszielen in der Wikipedia präsentieren dagegen Head und Eisenberg [Head und Eisenberg, 2010]. Dabei wurden College Studenten nach ihren Zielen bei der Benutzung der Wikipedia befragt.

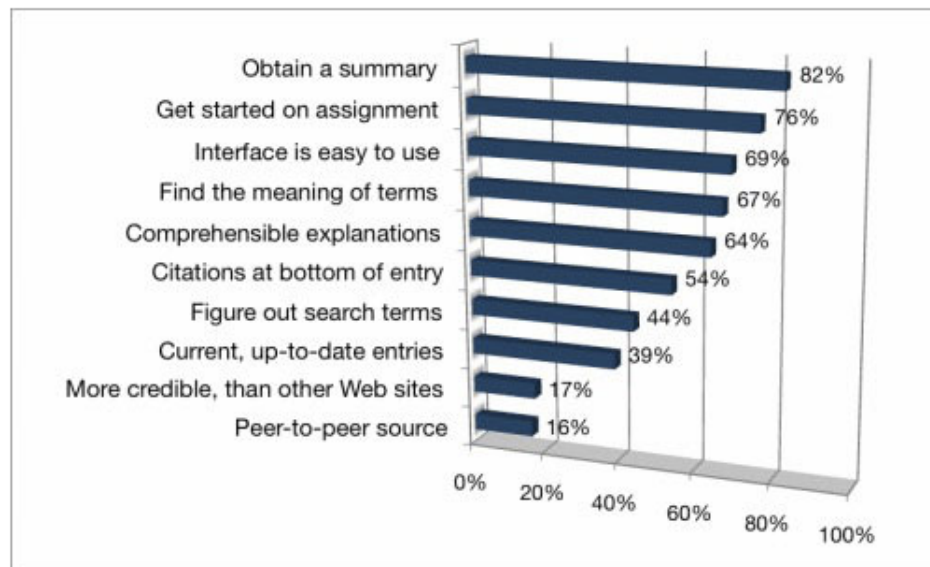


Abbildung 5-6 Head und Eisenberg, 2010, Ziele der Wikipedianutzung bei einer Umfrage bei College-Studenten 2008

Diese Auflistung der Ergebnisse enthält verschiedene Variablen, die mögliche Ziele der Wikipedia-Suche beschreiben. Zum Beispiel wird der Grund dafür genannt, die Wikipedia (und kein anderes System) auszuwählen („*More credible, than other websites*“, „*Interface is easy to use*“, siehe Abbildung 5-6) aber auch die tatsächlichen inhaltlichen Suchziele entsprechend der in dieser Arbeit vorgenommenen Definition wie: „*Obtain a summary*“, „*Figure out search terms*“ oder „*Citations at bottom of entry*“. Auch wenn diese Klassifikation schwieriger zu generalisieren ist (da sich bei der Auflistung die Variablen wie Motivation, inhaltliches Ziel und Argumente für die konkrete Wahl der Wikipedia überschneiden), so zeigt die Untersuchung von Head und Eisenberg, dass unterschiedliche Ziele mit Hilfe der Wikipedia zu erreichen versucht werden.

Auch Lim [Lim, 2009] führte eine ähnlich angelegte Umfrage durch und zeigt ebenfalls, dass der *Information Need* der College Studenten, die die Wikipedia benutzen, unterschiedlich ausfällt [Lim, 2009]. Nachfolgende Abbildung zeigt die Informationsbedürfnisse, deren Relevanz die Studenten jeweils auf einer Skala von 1-7 Punkten bewerteten. Sowohl Fakten Retrieval als auch ausführliche Informationsrecherche werden mit einer hohen durchschnittlichen Punktzahl genannt. Es sind also in der Wikipedia unterschiedliche Suchziele auszumachen.

Cognitive need	To look up a quick fact	5.85	1.329	$\alpha=0.824$ , mean: 5.16
	To browse information	4.72	1.804	
	To learn something I am not familiar with	5.19	1.677	
	To get more information on topics I want to learn more about	4.74	1.749	

Abbildung 5-7 Auszug aus der Studie von [Lim, 2009] zur Wikipedia Nutzung - cognitive needs

Hu et al. [Hu et al., 2009] erarbeiteten eine automatisierte Erkennung des *query intent* bei der Suche in der Wikipedia, dabei stand die rein *inhaltliche* Absicht des Nutzers im Vordergrund. Die Autoren trainierten einen *query intent predictor* in verschiedenen Domains mittels repräsentativer Anfragen und konnten so dann auch unbekannte Suchanfragen zuordnen. Der entwickelte Algorithmus basiert vor allem auf der Annahme, dass verlinkte Inhalte in der Wikipedia semantisch eng verwandte Konzepte sind. Die Vorhersagen funktionierten gut bei [Hu et al., 2009], allerdings ist der Schwerpunkt der Vorhersage der rein inhaltlichen Absicht zu vage für eine Zuordnung zu bestimmten Such-task-Klassen.

Weitere Arbeiten zur Verbesserung der Suchanfrage (*Query Refinement*, [Guo et al., 2008], *Query Recommendation* [He et al., 2009], *Query Expansion* [Fonseca et al., 2005] etc.) werden im Kapitel zum adaptiven Information Retrieval besprochen. Der Aspekt der Unterstützung von Suchanfragen z.B. durch die Berücksichtigung vorangegangener Suchanfragen bezieht sich auf inhaltliche Ziele einer Suche im Sinne von Topic/Domain. Die hier beschriebenen Variablen zur Definition eines Suchziels bleiben davon unberührt.

## 5.5 Unterscheidung nach der Motivation

Die zweite wesentliche Unterscheidung der Nutzungsszenarien in der Wikipedia findet nach der Motivation des Nutzers zur Suche statt. Nach Maslow [Maslow, 1943] gibt es für jede Handlung mehr als nur *eine* Motivation. Jeder Zustand eines Wesens ist durch irgendetwas motiviert und motiviert wiederum einen weiteren Zustand: „*Man is*



*a perpetually wanting animal*“ [Reid-Cunningham, 2008, S.58], entspricht in etwa auch [Wilson, 1999].

Der Antrieb und jede Motivation ist generell verbunden mit dem Wunsch nach Zufriedenheit. Deswegen postuliert Maslow, dass es wenig Sinn hat, eine Liste von konkreten Motivationsgründen zu erstellen, um eine Handlung zu erklären, sondern empfiehlt, dazu *Klassen* von Motivationsgründen zu verwenden. Entscheidend dabei ist, dass es nicht notwendig ist, immer jeweils die ganz genaue Situation zu erfahren (siehe [Maslow, 1943]), sondern eine formale, allgemeine Beschreibung ausreicht, um damit weiterzuarbeiten zu können. Allerdings muss jede konkrete Situation/Motivation anhand klarer Faktoren einer der Klassen zugeordnet werden können. Generell ist es aber einfacher, den Klassen bestimmte Eigenschaften und aus den Eigenschaften folgende Konsequenzen zuzuschreiben (z.B. für eine adaptive Informationspräsentation), als das jeweils am Einzelfall vornehmen zu müssen. Deshalb wird in dieser Arbeit das Prinzip der Klassen übernommen.

Weiterhin muss das jeweilige Umfeld, in dem diese Motivationsgründe greifen (hier: Suche in der Wikipedia) mit einbezogen werden. Durch die Einführung eines *nicht-informationsbezogenen Suchziels* (siehe dazu Kapitel 6.2.1) wird eine eher psychische oder physische (im Gegensatz zu der rein kognitiven) Motivation für die Suche ebenfalls berücksichtigt. Das entspricht den Erkenntnissen von [Wilson, 1999], der postuliert, dass das Entstehen eines Informationsbedürfnisses die Folge eines ursprünglicheren Bedürfnisses physischer, affektiver oder natürlich auch kognitiver Art ist. Erst daraus formuliert sich ein ASK [Belkin, 1980] oder aber z.B. ein nicht-informationsbezogenes Ziel wie Ablenkung, Unterhaltung, Zerstreuung etc.

Es wird hier also grundsätzlich in zwei Klassen unterschieden zwischen einem Informationsbedarf im Sinne des *anomalous state of knowledge* nach [Belkin, 1980] und damit als *work-based* bezeichneten Suchszenario und entsprechend den *non-worked-based* Szenarien. In den *work-based* Szenarien besteht der Bedarf, einen Ausnahmezustand des Wissens [Belkin, 1980] zu beseitigen, indem die dafür geeigneten Informationen gefunden werden. Die *non-work-based* Szenarien sind durch andere Gründe motiviert, wie im Folgenden gezeigt wird.

***Non-work-based Szenarien - die casual leisure-Theorie***

Im Fokus der dieser Theorie stehen nicht sog. *work-related tasks*, die klassisch dadurch ausgelöst werden, dass beim Nutzer ein konkretes Informationsbedürfnis besteht, sondern die Vermutung, dass viele Nutzer von Medien wie Twitter, Ebay, Amazon und Wikipedia zumindest teilweise ohne konkretes, informationsbezogenes Ziel browsen, suchen und navigieren [Elsweiler et al., 2011].

Wie Elsweiler et al. [Elsweiler et al., 2011] beschreiben, bezogen sich die meisten bisher durchgeführten Studien zum Suchverhalten vorwiegend auf einen Arbeitskontext bzw. zumindest auf Szenarien, in denen der Nutzer eine gewissen Informationslücke schließen will, oder aber die Untersuchungen fanden von vorherein unabhängig von dieser grundlegenden Unterscheidung statt (z.B. [Broder, 2002; Kang und Kim, 2003] etc.).

*Work-related tasks* standen also im Fokus der Forschung, wogegen die Suche in digitalen Informationssystemen aus mehr oder weniger ausschließlich hedonistischen Motiven auch erst mit den technologischen Fortschritten möglich und interessant wurde [Marchionini, 2006]. Um der Vermutung nachzugehen, dass sich Suchtasks in sog. *casual-leisure* Szenarien (z.B. nach [Stebbins, 2007] 8 Typen) von *work-related tasks* unterscheiden, führten [Elsweiler et al., 2011] zwei Studien durch. In der ersten Studie stand im Mittelpunkt der Untersuchungen die Motivation von Testpersonen dafür, den Fernseher anzuschalten bzw. fern zu sehen. Im Ergebnis zeigte sich eine Bandbreite von klar definierten Zielen bis hin zu überhaupt nicht mehr *work-related* Gründen („kill time“) [Elsweiler et al., 2011, S.9].

Es wurde auch deutlich, dass verschiedene weitere Faktoren über ein eigentliches Informationsbedürfnis die Motivationen der Nutzer beeinflussten:

User motivated	<i>Personal Interest, Knowledge or Lifestyle, Habits:</i>
	(a) "I am interested in food and learning new recipes"
	(b) "I like to keep up to date as it is related to my work"
Context motivated	(c) "I tend to watch soaps in the evenings"
	<i>Mood or State:</i>
	(d) "I want to be thrilled / entertained / distracted"
	(e) "I'm tired / bored / curious / frustrated"
	<i>Time Related:</i>
	(f) "It is ...8.15pm, ...after work, ... the weekend"
	(g) "I have 30 mins to kill"
	<i>Socially motivated:</i>
	(h) "My mum was watching it..."
	(i) "...something that my dad and I can watch together"
Planning	<i>Programme Related:</i>
	(j) "I saw an advert for computer games" "so that I can plan my day / week/ weekend"

Abbildung 5-8 Auszug aus der Diary Study von [Elsweiler et al., 2011, S. 10]

Weiterhin fanden die Autoren heraus, dass gerade in den Situationen, in denen die Motivation hauptsächlich oder ausschließlich hedonistischer Natur war, die Zufriedenheit mit dem Ergebnis zum größten Teil nur von der Befriedigung des – in diesem Fall nicht konkret informationsbezogenen (im Sinne des ASK von [Belkin, 1980]) – Bedürfnisses abhängt (z.B. Langweile beseitigen, „auf andere Gedanken kommen“ etc.), und nicht so sehr vom Auffinden einer bestimmten Information. Das ist ein nachvollziehbarer Beweis dafür, dass der Erfolg bzw. die für den Erfolg notwendigen Suchergebnisse immer subjektiv und in Abhängigkeit des auslösenden Bedürfnisses beurteilt werden. Die Herausforderung besteht nun darin, diese Befriedigung und Bewertungen konkret in Abhängigkeit von der Motivation für die Suche in der Wikipedia zu untersuchen.

Auch in der weiterführenden Studie der Autoren, bei der durch Aggregation von *Tweets* der Mikroblogging Plattform Twitter die Motivationen der Nutzer untersucht wurden, zeigte sich der selbe Trend – allerdings generalisierter: Wiederum fiel auf, dass in vielen Fällen *Searching*, *Browsing* und *Surfing* im Freizeitbereich stattfand (Siehe Abbildung 5-9).

- 1) ... I'm not even *\*doing\** anything useful... just browsing eBay aimlessly...
- 2) to do list today: browse the Internet until fasting break time..
- 3) ... just got done eating dinner and my family is watching the football. Rather browse on the laptop
- 4) I'm at the dolphin mall. Just browsing.
- 5) Cannot sleep...la di da...browsing iPhone apps.

Table 4

Example tweets where the browsing activity is need-less.

Abbildung 5-9 Auszug aus der Twitter-Studie von [Elsweiler et al., 2011, S.12]

Zwar wird aus dieser Studie nicht direkt ersichtlich, wie die in den Tweets benannte Absicht (z.B. „...rather browse on the laptop“) jeweils inhaltlich umgesetzt wurde, aber die Erkenntnis, dass das Stöbern in digitalen Ressourcen nicht grundsätzlich *work-related* ist und dass das Ziel nicht immer ist, eine mehr oder weniger genau definierte Informationslücke zu schließen, zeigen die Untersuchungen von [Elsweiler et al., 2011-1; Elsweiler et al., 2011 und Elsweiler und Wilson, 2010] deutlich.

Als ein Fazit definieren die Autoren [Elsweiler et al., 2011-1] weiterhin vier wesentliche beobachtbare Unterschiede im Verhalten der Nutzer:

1. *Casual-Leisure Information Behaviours tasks are often motivated by being in or wanting to achieve a particular mood or state. Tasks often relate at a higher level to quality of life and health of the individual.*
2. *In Casual-Leisure situations, therefore, the finding of information is often of secondary importance to the experience of finding.*
3. *Casual-Leisure needs are frequently associated with very under-defined or absent information needs. We hypothesize that transient or temporally-defined needs are created in order to facilitate meeting the casual need, but our studies have, so far, not directly studied searching behavior*
4. *The success of casual-leisure behaviour is often not dependent on actually finding any information or results. Even if notable information needs exist, the low importance of the task means that users satisfice or can achieve their hedonistic needs without them.*

[Elsweiler et al., 2011-1, S.14]

Auch bei rein hedonistischer Motivation kann sowohl das Bedürfnis bestehen, sich über ein Thema ausführlich zu informieren (z.B. *knowledge acquisition* [Marchionini, 2006]) als auch ein zielloses Stöbern stattfinden [Adafre und Rijke, 2006]. Dabei handelt es sich um eine vom Interesse oder dem Zufall geleitete Suche (das ist bei strenger

Sichtweise im Prinzip keine Fragestellung des „klassischen“ Information Retrieval, für die Fragestellung dieser Arbeit aber deshalb relevant, da die Präferenzen nicht nur auf das „klassische“ IR beschränkt untersucht werden sollen).

Diese in diesem Kapitel beschriebene Motivation zur Suche steht der *work-based* Suche gegenüber. Somit ist hier festzuhalten, dass erstens unterschiedliche Situationen vorkommen, in denen ein Suchimpuls entsteht, zweitens verschiedene Inhaltliche Ziele möglich sind und drittens verschiedene Motivationen eine Suche auslösen. Damit wurde gezeigt, dass im Information Retrieval grundsätzlich unterschiedliche und voneinander abgrenzbare Nutzungsszenarien entstehen.

## 5.6 Unterschiedliche Szenarien – weitere Einflüsse

Neben den hier genannten Unterscheidungen der Nutzerszenarien beim Information Retrieval existieren weitere Faktoren, die das Verhalten und die Vorstellung der Zielerreichung der Nutzer lenken.

### 5.6.1 Durchsuchter Wissensbereich

In der Literatur finden sich Hinweise, dass der durchsuchte Wissensbereich [Yue et al., 2013; Han et al., 2013] bzw. der Detaillierungsgrad der Inhalte eines Dokumentes [Yan et al., 2011] ebenfalls eine Auswirkung auf das Suchverhalten der Nutzer hat. Auch [Calderon-Benavides et al., 2010, S.3] stellten einen Einfluss des *topics* (z.B. *adult & Sex, Arts & Culture, Beauty & Style, Cars & Transportation, Computers & Internet, Education, Entertainment, Music & Games, Finance, Food& Drink, Health, Home & Garden*) fest. Diese Erkenntnisse beziehen sich allerdings wiederum auf die Websuche wie auch die entsprechend daraus abgeleiteten Mechanismen zur Verbesserung der Suchergebnisse. Es fehlen Hinweise, wie sich dieser Einflussfaktor konkret themenbezogen bei der Suche in der Wikipedia auswirkt. Kelly und Teevan [Kelly und Teevan, 2003] weisen ebenfalls darauf hin, dass der Gegenstand eines Textes einen Einfluss auf die Lesezeit der Nutzer hat.

### 5.6.2 Art des Inhaltstyps – Betonung des Medientyps Bild

Es liegen keine konkreten Hinweise darüber vor, welche Inhaltstypen abhängig vom Nutzungsszenario bevorzugt werden. Bei Bildern allerdings geht aus der aktuellen Forschungslage eindeutig hervor, dass Mehrwert und Lernförderlichkeit von Bildern auch stark von der Rezeption, den Eigenschaften und der Rolle eines Bildes im Gesamtzusammenhang abhängt (u.a. [Nöth, 2008; Weidenmann, 1991], siehe dazu Kapitel 1). Die Zusammenhänge bevorzugter Inhaltstypen und vor allem Bildern im Bereich der Informationssuche in der Wikipedia sind aber ungeklärt.

### 5.6.3 Erfahrung und Vorkenntnis

Im Bereich des E-Learning wurde gezeigt, dass eine Berücksichtigung des Erfahrungsstandes des Nutzers verbesserte Ergebnisse lieferte [Trumbley et al., 1994].

White et al. [White et al., 2009] führten eine Langzeitstudie durch, in der sie untersuchten, in wie weit sich Themen-Experten und Novizen in ihrem Suchverhalten unterscheiden. Dabei konzentrierten sie sich auf die jeweiligen Vorkenntnisse in einer bestimmten Domain (hier: Medizin, Finanzen, Recht und Informatik). Es gelang den Autoren, Experten anhand ihrer Suche deutlich von Themen-Novizen zu unterscheiden [White et al., 2009]. Dabei wurden allerdings vor allem die besuchten Inhalte und Seiten im Web (hinsichtlich derer generellen Relevanz für die jeweiligen Fragestellungen) herangezogen und nicht das Suchverhalten der Nutzer an sich. Dieser Aspekt bleibt offen.

Kelly et al. [Kelly et al., 2002] zeigen, dass aus dem Interaktionsverhalten der Nutzer bis zu einem gewissen Grad auf deren Vertrautheit mit einem Thema zurückgeschlossen werden kann – z.B. durch die Dauer der Lesezeit, die sich bei großer Themenerfahrung deutlich verringert. Im Zusammenhang mit dieser Arbeit ist von Interesse, ob das auch bei der Recherche in der Wikipedia gilt und so aus der Interaktion themenerfahrene Nutzer erkannt werden können.

Marchionini [Marchionini, 1989] betont ebenfalls, dass Erfahrung und Vorwissen des Nutzers die Suchtaktiken und die Interaktion mit dem System während der Suche beeinflussen. So sind zum Beispiel bei Novizen die Suchstrategien einfacher und die Navigation im Informationssystem eher durch kurze und direkte Pfade gekennzeichnet.

Einen weiteren Ansatz, um den Einfluss des *domain knowledge* zu untersuchen, verfolgte [Wildemuth, 2004]. Die Autorin ließ Medizinstudenten in einer Datenbank mit In-

halten der Domain Mikrobiologie nach jeweils benötigten Informationen suchen – und zwar über einen Zeitraum von neun Monaten. So konnte sie beobachten, dass sich, je vertrauter die Studenten mit dem Thema wurden, auch ihre Suchstrategien änderten. Dabei mussten mit Hilfe der Datenbank immer neue Probleme gelöst werden, dennoch wurden die Studenten immer erfahrener und konnten schneller ihren Themenbereich eingrenzen, lasen weniger etc. Dieser Ansatz von [Wildemuth, 2004] zeigt, dass die Themenerfahrung einen Einfluss auf das Suchverhalten der Nutzer hat und somit im Sinne eines Nutzungsszenarios berücksichtigenswert ist.

#### 5.6.4 Persönliche Faktoren

Marchionini [Marchionini, 1989] macht in einer Studie mit Grundschulkindern und Erwachsenen deutlich, dass auch das Alter eine Rolle dabei spielt, wie der Nutzer bei der Suche vorgeht. Auch Jochmann et al. [Jochmann et al., 2010] zeigen, dass Kinder anders suchen als Erwachsene.

Es können weiterhin auch eventuelle Einschränkungen einer Person, wie z.B. eine Les- und Rechtschreibschwäche oder Lernschwierigkeiten, eingeschränktes Sehvermögen etc. berücksichtigt werden. Über die Informationsaufnahme und -wahrnehmung von Personen mit kognitiven oder physischen Einschränkungen sind wichtige Hinweise bei [Dahm, 2006] zu finden.

#### 5.6.5 Der Lerntyp

Einen weiteren Einfluss auf die Suche und die Rezeption der Inhalte ist die Frage, welchem Lerntyp der jeweilige Nutzer *generell* angehört. In der pädagogischen Forschung gibt es verschiedene Lerntypen [Felder, 1993]. Nach [Felder, 1993] spielt das in der Präsenzlehre eine große Rolle: Wenn der Unterrichtsstil nicht dem Lerntyp entspricht (z.B. bevorzugtes Lernen durch Visualisieren oder Wiederholen), sind die Lernerfolge geringer, als wenn die Methoden entsprechend abgestimmt sind. Es lässt sich daraus die Vermutung ableiten, dass der Lerntyp bei der digitalen Informationsvermittlung im autodidaktischen Szenario ebenso eine Rolle spielt wie in der Präsenzlehre.

Lerntypen setzen sich aus verschiedenen Vorlieben zusammen, nach [Felder, 1993]

- 1) bevorzugter Informationstyp: *sensorisch* (Geräusche, etwas Sehen etc.) oder *intuitiv* (Erinnerungen, Gedanken, Ideen etc.)

- 2) bevorzugte Modalität: *visuell* (Bilder, Diagramme, Graphen etc.) oder *verbal* (Sound, geschriebene und gesprochene Sprache und Wörter)
- 3) Art der Informationsorganisation: *induktiv* (vom Anwendungsfall ausgehend nach allgemeinen Prinzipien suchend) oder *deduktiv* (allgemeine Prinzipien liegen vor und der Anwendungsfall wird untersucht)
- 4) bevorzugte Informationsverarbeitung: *aktiv* (physischer Einsatz oder Diskussion) oder *reflektierend* (durch Überlegen)
- 5) Art des Verstehens-Prozess: *sequentiell* (logische Abfolge einzelner Schritte) oder *global* (holistisch und/oder in großen Sprüngen)

### 5.6.6 Schwierigkeit und Komplexität des Tasks

Studien haben weiter gezeigt, dass die Komplexität der Fragestellung entweder direkt [Liu et al., 2012] oder durch die jeweiligen Emotionen, die z.B. ein hoher Schwierigkeitsgrad auslöst [Moshfeghi und Jose, 2013] das Verhalten der Nutzer und deren Reaktionen auf das System beeinflussen (u.a. [Poddar und Ruthven, 2010]). Liu et al. [Liu et al., 2012] zeigten deutliche Unterschiede im Verhalten der Nutzer abhängig davon, ob es sich um einen einfachen oder einen schwierigen Suchtask handelte. Sowohl zu Beginn der Suche als auch während der Suche und über den kompletten Suchtask gesehen unterschieden sich die Suchstrategien der Nutzer in Abhängigkeit von dem Schwierigkeitsgrad der Aufgabe [Liu et al., 2012].

In den Untersuchungen stellen die Autoren auch fest, dass bei bestimmten Tasktypen die Bewertung des Schwierigkeitslevels durch den Benutzer auch von weiteren Faktoren, wie zum Beispiel der verfügbaren Zeit abhing. Demzufolge wurden Tasks vor und nach der Ausführung als unterschiedlich schwierig eingestuft [Liu et al., 2012]. Mit einem auf Basis der Ergebnisse entwickelten Vorhersagemodell verfolgen Liu et al. [Liu et al., 2012] ebenfalls das Ziel, Suchergebnisse entsprechend dem Nutzerverhalten adaptiv zu gestalten.

### 5.6.7 Einfluss von Emotionen

In zweierlei Hinsicht spielt auch der Einfluss von Emotionen eine wesentliche Rolle. Zum einen als Größe in den *casual-leisure* Szenarien, bei denen die Verbesserung der emotionalen Lage als zentrales Ziel auftreten kann [Elsweiler et al., 2011]. Zum anderen zeigen [Axelrod, 2004; Moshfeghi, 2009 so wie auch Nahl und Bilal, 2007] auch, dass



affektive (und nicht nur kognitive) Faktoren die Bedürfnisse eines Nutzers bei der Suche in digitalen Systemen beeinflussen.

Studien haben gezeigt, dass Emotionen auch im *Information Seeking Process* eine Rolle spielen [Nahl und Bilal, 2007; Axelrod, 2004; Poddar und Ruthven, 2010; Norman 2005; Arapakis et al., 2008 und Moshfeghi und Jose, 2013]. Gerade in Situationen, in denen Langeweile, Anspannung oder Stress herrschen, ist es damit möglich, dass die Präferenzen für Inhalte unterschiedlich ausfallen.

Einerseits kann dies aus der grundsätzlichen Emotionsforschung und dem Wissen darüber abgeleitet werden [Scherer, 1990], dass Emotionen immer eine Rolle spielen, wenn Menschen Entscheidungen treffen [De Sousa, 1990]. Andererseits zeigen gezielte Arbeiten, dass während des Suchprozesses Wechsel in den Emotionen zu erkennen sind die sich auch abhängig von den Inhalten und dem Fortschritt der Suche zeigen [Kuhlthau, 1991] und [Moshfeghi und Jose, 2013].

Emotionale Erfahrungen können bei der Interaktion von Menschen mit Maschinen gemessen werden [Lopatovska, 2009]. Da sie das Verhalten der Menschen beeinflussen [De Sousa, 1990; Scherer, 1990], ist somit auch eine Auswirkung auf das Suchverhalten gegeben.

Im Rahmen einer Standard-Mensch-Maschine-Schnittstelle bleibt dies aber in der Regel unberücksichtigt – zumindest auf Seiten der Maschine [Picard, 1998]. Der emotionale Zustand des Nutzers kann sich aber verändern, entweder als Reaktion auf die Interaktion (als gewollt herbeigeführtes Ziel oder ungewollt) oder durch andere, äußere oder innere Einflüsse [Nahl und Bilal, 2007]. Nahl und Bilal beschreiben den Prozess der Informationsaufnahme und -verarbeitung als Zusammenspiel dreier unterschiedlicher Wahrnehmungsebenen: der sensomotorischen, der kognitiven und der affektiven. Die affektive Komponente ist also ebenfalls entscheidend (im Gegensatz zu einer reinen Fokussierung auf den Recall und die Precision entsprechend dem kognitiven Defizit).

Picard [Picard, 1998] zeigt in einer Studie exemplarisch, welches Potential dahinter steckt, wenn Systeme auf bestimmte affektive Zustände des Nutzers reagieren können, vor allem auf Frustration oder Ärger. Picard verwendet dazu zwei Wege, die Emotionen des Nutzers zu erfassen: zum einen wird versucht, anhand der Interaktion auto-

matisch Rückschlüsse zu ziehen, zum anderen geben die Nutzer selber Auskunft über ihren emotionalen Zustand. [Picard, 1998] beschreibt Möglichkeiten, diese Auskünfte in der Mensch-Maschine-Interaktion zu berücksichtigen, vor allem im Bereich *Entertainment* und *Education*. Dazu wurde ein System implementiert, mit dem gezeigt werden konnte, dass es durchaus möglich ist, zum Beispiel durch Anpassung der Geschwindigkeit auf z.B. Frustration des Nutzers zu reagieren. Diese Erkenntnis kann auch für die Informationssuche übernommen werden, da erstens bestimmte Emotionen aus der Interaktion erkennbar sind (hierzu auch [Feild et al., 2010]) und zweitens eine Berücksichtigung dieser die Zufriedenheit der Nutzer steigern kann [Picard, 1998].

Poddar und Ruthven [Poddar und Ruthven, 2010] beschäftigen sich ebenfalls mit dem Zusammenhang der Emotionen des Nutzers und bestimmten weiteren Variablen eines Suchtasks. Dazu wurden die VPs gebeten, verschiedenen Suchaufgaben zu lösen, die vom *fact finding* bis hin zum *exploratory learning* in Komplexität, Umfang und persönlicher Gestaltungsfreiheit variierten. Es wurden von jeder Testperson Daten wie Vorwissen, Erfahrung bei der Informationssuche und persönliche Interessen abgefragt.

Zusätzlich wurden die VPs gebeten, ihre Gedanken und Erfahrungen während der Suche laut mitzusprechen (*think aloud*). Ein Beobachter versuchte gleichzeitig, non-verbale Ausdrücke mit zu protokollieren, die Hinweise für affektive Reaktionen sein könnten. Mit Hilfe dieses Versuchsaufbau wurde angestrebt, herauszufinden, ob die Teilnehmer bestimmte emotionale Muster im Zusammenhang mit den einzelnen Tasktypen durchliefen, ob die Emotionen unterschiedlich ausfielen, je nachdem ob die Nutzer einen künstlich konstruierten Task lösen mussten oder nach eigenem Interesse explorativ vorgehen durften. Außerdem war hier von Bedeutung, welche Faktoren der Interaktion am wichtigsten sind, um die Emotionen vorhersagen und erkennen zu können. Die Autoren fanden heraus, dass vor allem bei den eigenen (nicht vorgegeben) Search Tasks die Emotionen vor und nach der Suche besser waren als bei den konstruierten Aufgaben [Poddar und Ruthven, 2010]. Damit wurde hier ein Einfluss des Aufgabentyps auf die Emotionen des Nutzers dargelegt.

Arapakis et al. [Arapakis et al., 2008] untersuchten ebenfalls den Einfluss von unterschiedlichen Suchtasks auf die emotionale Lage der Personen. Auch sie konnten zei-

gen, dass zum Beispiel die Komplexität einer Aufgabe einen Einfluss auf die Emotionen der Nutzer hat. Weiter beschreiben Arapakis et al., dass auch der Schwierigkeitsgrad der Suche die Emotionen der Personen von einem eher positiven zu einem eher negativen Bereich bewegten. In der Konsequenz dazu erkannten die Autoren weiter, dass ein allgemeines, komplexes Zusammenspiel zwischen der Suche, den Emotionen der Probanden, deren kognitiven und auch physischen Prozessen und letztendlich wiederum mit den angewandten Suchstrategien vorliegt [Arapakis et al., 2008].

Die beschriebenen Studien legen dar, dass Emotionen sowohl als Zielgröße als auch als Einflussgröße auf das Suchverhalten in der Interaktion mit Systemen im Allgemeinen und dem Information Retrieval im Speziellen eine Rolle spielen. Es ergibt sich die Frage, ob dies auch auf die Suche in der Wikipedia zutrifft.

Eine weitere Problematik beim Thema der Emotionen ist die Komplexität sowohl bei der Beschreibung als auch der Erfassung selbiger, vor allem im Bereich der Mensch-Maschine-Interaktion. Was sind Emotionen eigentlich, wie beeinflussen sie den Menschen und welche Emotionsmodelle sind für die Mensch-Maschine-Interaktion relevant? Zentral ist auch die Frage, wie und in wie weit emotionale Zustände in der Kommunikation zwischen Nutzer und Computer im Allgemeinen und bei der Suche in der Wikipedia im Speziellen erfasst und berücksichtigt werden können. Diesen Fragestellungen wird im empirischen Teil dieser Arbeit nachgegangen (siehe ab Kapitel 11.1.1).



## 6 Präzisierung in einem Modell

Auf Basis der Literatur wurde dargestellt, dass verschiedene Nutzungsszenarien im Information Retrieval existieren und nach verschiedenen Aspekten unterscheidbar sind. Dabei spielen Suchauslöser, inhaltliche Vorstellungen und weitere Faktoren aus dem Umfeld und Kontext des Nutzers eine Rolle, die umso heterogener sein können, als dass gezeigt wurde, dass die Suche in digitalen Ressourcen fester Bestandteil unseres Alltags geworden ist.

Die verschiedenen Einflussfaktoren werden in einem Modell zusammengefasst, präzisiert und auf den Forschungsschwerpunkt dieser Arbeit, der Suche in der Wikipedia, bezogen.

### 6.1 Entwicklung eines Modells

Ein Modell ist immer zweckabhängig. Es bildet damit die Realität nicht vollständig ab, sondern abstrahiert und enthält die für den Anwendungsfall wichtigsten Elemente. Damit unternimmt es automatisch eine Verkürzung der Realität. Deshalb muss ein Modell anwendungsspezifisch entwickelt und/oder aus vergleichbaren Anwendungsfällen adaptiert werden.

Die vorhandenen Modelle zum Verhalten der Nutzer bei der Suche in digitalen Informationssystemen (siehe [Ingwersen und Järvelin, 2005]) sind zahlreich aber in der Regel relativ allgemein. Es ist schwierig, mit diesen konkrete, operationalisierte Einflüsse für die Informationssuche in der Wikipedia darzustellen. Der Kontext wird nicht im Detail abgebildet, wenn gleich er auch berücksichtigt wird und als wesentlicher Teil der Interaktion zwischen Nutzer und System mit einfließt (siehe z.B. [Wilson, 1999]). Deshalb ist es ein wichtiger Schritt, bisherige Modellteile wie „persönliche Faktoren“ oder „Informationen entsprechend dem Suchziel“ bzw. „Einflüsse auf kognitiver und affektiver Ebene“ zu konkretisieren.

Allgemein beziehen sich die vorhandenen Modelle des *Information Seeking* vor allem auf eher systematische Suchprozesse (z.B. [Kuhlthau, 1991]). Auf eine Suche, die *kein* inhaltliches Ziel hat, treffen deshalb verschiedene Abschnitte des Suchprozesses unter

Umständen nicht zu. Im hier entwickelten Modell wird deshalb zusätzlich das nicht-informationsbezogene (also nicht auf das Auffinden einer Information gerichtete) Suchziel berücksichtigt.

Für die Entwicklung des vorliegenden Modells sind die Nutzungsszenarien der Wikipedia so wie Rezeption von Information und das Suchverhalten der Nutzer relevant. Durch die Beschränkung auf die Nutzung der Wikipedia fallen weitere, anderenfalls eventuell relevante Aspekte, wie etwa die Auswahl des Informationssystems, weg.

Im hier entwickelten Modell wurden folgende Faktoren berücksichtigt:

1. Einflussfaktoren aus der besprochenen Literatur (siehe entsprechende Kapitel)
2. Faktoren, die speziell die Suche in der Wikipedia berücksichtigen (hier vor allem vorhandene inhaltliche Strukturen und Konzepte)
3. Abstrakte Konzepte, wie die „*task completion*“, die sich aus logischen Schlussfolgerungen zu Überlegungen wie dem *cognitive viewpoint* (siehe [Ingerwersen, Jarvelin, 2005]) und den existierenden Modellen des *Information Seeking* ergeben und der Relevanz, die sich durch die Dauer ergibt, mit der der Nutzer Inhalte betrachtet hat (siehe dazu auch [Claypool et al., 2001; White et al., 2005; Buscher et al., 2008]).

Die Darstellung und Verwendung der Einflussfaktoren und des Nutzerverhaltens im Modell (z.B. feste Wertebereiche für die Variablen) wurden vor allem der Benutzbarkeit und dem intendierten Zweck (Übersicht und praktische Anwendung) entsprechend gestaltet.

Das Modell beschreibt die *generelle Sichtweise* auf die untersuchte Situation und fasst die in der Literatur genannten und für die Hypothese dieser Arbeit entscheidenden Faktoren zusammen.

Ingwersen und Järvelin formulieren sieben Kriterien, anhand derer der Nutzen und die Güte eines Modells (bzw. mehrerer Modelle im Vergleich) evaluiert werden können.

[Ingwersen und Järvelin, 2005, S.14]:

[ii] When two competing conceptual models are compared the following criteria may be applied to judge their merits:

- *Simplicity: simpler is better other things being equal.*
- *Accuracy: accuracy and explicitness in concepts is desirable.*
- *Scope: a broader scope is better because it subsumes narrower ones, other things being equal.*
- *Systematic power: the ability to organize concepts, relationships and data in meaningful systematic ways is desirable.*
- *Explanatory power: the ability to explain phenomena reliably and to predict them is desirable.*
- *Validity: the ability to provide valid representations and findings is desirable.*
- *Fruitfulness: the ability of a model to suggest problems for solving and hypotheses for testing is desirable.*

Die Berücksichtigung dieser Kriterien für ein gutes Modell führen dazu, dass die Erklärungen zu den Variablen und deren Wertebereichen in dieser Arbeit vom allgemeinen Modell getrennt dargestellt werden.

Die hier konkret angegebenen Wertebereiche verbessern die Genauigkeit in der Anwendung. Die Darstellung des kompletten Modells beschreibt die einzelnen Stufen und Bereiche in ihrem Zusammenhang.

Auf Basis dieses Modells wurden Studien im Forschungsgebiet der Wikipedia-Suche geplant und durchgeführt [siehe auch Schubart, 2013; Vogel, 2013]. Durch einheitliche Verwendung und Planung können die Versuchsergebnisse verglichen und ergänzend verwendet werden. Da in den in der Fachliteratur vorliegenden Studien jeweils unterschiedliche Metriken zur Erfassung der Variablen verwendet wurden, sind deren Ergebnisse nicht direkt vergleichbar. Durch einheitliche Anwendung von Erfassungsska-

len können auch unabhängig voneinander durchgeführte Studien besser gemeinsam interpretiert werden.

## 6.2 Das Modell

Ziel des Modells ist es, die verschiedenen Erkenntnisse aus der Literatur zu komprimieren und in eine einheitliche formale Form zu überführen. Auf dieser Basis wurde das Modell erstellt. Im Folgenden werden die verschiedenen Modellbereiche einzeln erklärt, eine schematische Darstellung des Modells erfolgt im Anschluss.

Für die Stärke der Ausprägung eines Merkmals wird jeweils eine sechsstufige Likertskala verwendet. Es gibt keinen negativ-Bereich, sondern die niedrigste Ausprägung ist jeweils 0. Der Vorteil einer metrisch skalierten Erfassung der Ausprägung (im Gegensatz zu lediglich ordinalen Angaben wie „schwach“, „mittel“ und „stark“) ist, dass auf die erfassten Daten mathematische Operationen zur Analyse angewendet werden können. Eine detaillierte Aussage über den Kontext erhält man in *Kombination* der verschiedenen Variablen. Für die Variablen, die einen qualitativen Wertebereich haben, wurden die jeweils erstellten Klassifikationen verwendet.



### 6.2.1 Der Nutzer und sein Task

Der Nutzer und sein Informationsbedürfnis werden als Einheit formuliert, da die Gedanken und Ziele eines Suchbedürfnisses gleichermaßen beidem zugeordnet werden kann.

Modellierung des Nutzers und des Informationsbedürfnisses:

Tabelle 6-1 Modell des Nutzers und des Informationsbedürfnisses

#	Name	Wertebereich
V <sub>1</sub>	Motivation	M {workbased: 0 ...5}
V <sub>2</sub>	Suchziel	S <sub>1</sub> definiert= {0...5} S <sub>2</sub> endlich = {0...5} S <sub>3</sub> nicht informationsbezogenes Suchziel = {0...5}
V <sub>3</sub>	Stimmung	S <sub>t</sub> = {hier: SAM}
V <sub>4</sub>	Zufriedenheit	Z <sub>1</sub> = [Zufriedenheit mit dem Suchergebnis] {0...5} Z <sub>2</sub> = [Zufriedenheit mit dem Verlauf] {0...5}
V <sub>5</sub>	Sucherfolg	E <sub>S1+S2</sub> = [gefundene Information] {0...5} E <sub>S3</sub> = [Erfüllung des nibS] {0...5}
V <sub>6</sub>	Vorkenntnis	V <sub>Thema</sub> = {0...5} V <sub>WP-Recherche</sub> = {0...5}
V <sub>L</sub>	persönliche Präferenzen	Lerntyp nach [Felder, 1993]
V <sub>D</sub>	Demografische Faktoren	V <sub>D1</sub> = Alter V <sub>D2</sub> = Geschlecht V <sub>Dx</sub> = weitere Faktoren (Bildungsgrad etc.)
V <sub>C</sub>	Komplexität	V <sub>C</sub> = {1...5} [nach [Byström und Järvelin, 1995]

#### Beschreibung der Variablen im Modell:

V<sub>1</sub>: Variable eins beschreibt, in wie weit die Motivation zur Suche *work-based* ist. Konkret bedeutet das, ob der Nutzer ein Lernziel oder einen Wissenszustand zu einem bestimmten Zweck erreichen will. Diese Variable sagt noch nichts über die Suchstrategie oder den Umfang der gesuchten Informationen aus. Sie beschreibt, wie der Suchimpuls motiviert ist. Borlund beschreibt das als: „...a cognitive state which creates an information need“ und „... when one becomes aware of a mental state of current incompleteness“ [Borlund

und Ingwersen, 1997, S.8]. Ein nicht oder wenig *work-based* motivierter Task nach Borlund fällt in die Klasse der *casual-leisure tasks*.

V<sub>2</sub>: (S<sub>1</sub>+S<sub>2</sub>) beschreiben den Charakter des Suchziels selbst. Hier werden die Aspekte von [Rose und Levinson, 2004] aufgegriffen, die sich auf die Endlichkeit und die Definiertheit des inhaltlichen Suchziels beziehen. Endlichkeit bedeutet in diesem Fall, dass die *Vorstellung* des Nutzers vom inhaltlichen Suchziel endlich ist.

Belkin [Belkin, 1980] beschreibt 1980 das Konzept der „*non-specifiability of need*“ und die Tatsache, dass der Nutzer unter Umständen nicht genau weiß was er sucht, also auf kognitiver Ebene das Ziel nicht genau spezifiziert werden kann („Ich möchte irgendwo in den Urlaub fahren, wo es oft sonnig ist, aber nicht zu weit weg“). Er erklärt weiterhin, dass die „*non-specifiability of need*“ auch ein sprachliches Problem sein kann, sprich, der Nutzer in dem Moment der Anfrage nicht in der Lage ist, die Query tatsächlich seinem Informationsbedürfnis passend zu formulieren. Das weicht die Definiertheit des Suchziels für das System zunächst ebenfalls auf, kann sich aber im Laufe der Suche klären. Danach würde sich dann auch S<sub>1</sub> entsprechend anpassen.

Aus zwei Gründen wird hier die Variable des inhaltliche Ziels in Untervariablen aufgeteilt: Es besteht die Vermutung, dass sich die konkreten Suchziele und damit die aktuellen Suchstrategien während einer (z.B. in den empirischen Erhebungen als „ein Task“ definierten) Suchaufgabe ändern, und das lässt sich mit einem Verändern der Werte der Variablen S<sub>1</sub> – S<sub>3</sub> beschreiben. Sucht ein Nutzer nach möglichst vielen Informationen über das Spiel „Lacrosse“, wie in der ersten Studie zur Rezeption der Information in der Wikipedia dieser Arbeit, dann ist das Suchziel zunächst weniger endlich und auch nur mittelstark definiert (also wäre S<sub>1</sub> = 3 und S<sub>2</sub> = 2), außerdem verfolgt er zunächst ausschließlich das Ziel, sich über die Sportart so gut wie möglich zu einem bestimmten Zweck zu informieren, somit ergibt sich für S<sub>3</sub> = 1 oder 0. Während der Bearbeitung der Suchaufgabe könnte aber die Frage entstehen, seit wann dieses Spiel eigentlich in Deutschland gespielt wird. Es würden sich ändern S<sub>1</sub> = 5 und S<sub>2</sub> = 5, S<sub>3</sub> bliebe gleich. Gegen Ende der Bemühungen, so viele Informationen wie möglich über das Spiel zu akquirieren, ist vorstellbar, dass der Nutzer durch einen Hyperlink auf ein

anderes Thema stößt, das mit der initiativen Suchaufgaben nichts mehr zu tun hat, ihn aber zum weiterlesen animiert. Ohne bestimmten Informationsbedarf und nur aus reinem Interesse werden nun die Informationen konsumiert, wobei sich die einzelnen Variablen etwa auf Werte wie  $S_1 = 0$ ,  $S_2 = 2$  und  $S_3 = 5$  ändern.

Das entspricht auch den Überlegungen von Bates [Bates, 1989]: Eine einzelne Suchanfrage gibt es so gesehen nicht, sondern vielmehr eine Serie aus kleinen Einzelanfragen (die sogesehen in sich durchaus komplette Suchanfragen sind, wie oben beschrieben), die immer genau dem jeweiligen *information need* entsprechen. Sobald der Nutzer eine Information gefunden hat wird er dementsprechend reagieren und entweder weitere Ideen haben, die er in Anfragen umsetzt, oder bei ausreichender Information die Suche beenden.

Darüber hinaus müssen zweitens die Variablen  $S_1$  und  $S_2$  nicht miteinander korrelieren. Der Grad der Definiertheit eines Suchziels beeinflusst nicht zwingend die Endlichkeit und umgekehrt.

### ***Das Nicht-Informationsbezogene Suchziel $S_3$***

Um das „ziellose browsen“ (siehe dazu [Elsweiler und Wilson, 2010; Adafre und Rijke, 2006]) von der Suche nach einer bestimmten (endlich oder nicht endlich, definierten oder undefinierten) Information zu unterscheiden, wird das „nicht-informationsbezogene Suchziel“ eingeführt (*nibs*). Dieses liegt dann vor, wenn der Nutzer keinen *work-based task*<sup>12</sup> durchführt. Wichtig ist, hierbei zu betonen, dass es nicht zutrifft, dass der Nutzer dann kein Suchziel hat. Allerdings ist die *task completion* nichtzwingend messbar durch die gefundene Information bzw. den Grad der Befriedigung des Eingangs vorhandenen Informationsbedürfnisses, sondern dadurch, dass z.B. eine andere Stimmung erreicht wurde, dass der Nutzer eine Wartezeit überbrückt hat, Ideen gesammelt hat oder einfach nur abgelenkt war.  $S_3$  beschreibt also ein Suchziel, das nicht auf das Auffinden einer bestimmten Information gerichtet ist.

**V<sub>3</sub>:** Die Stimmung beschreibt den emotionalen Zustand des Nutzers zu Beginn der Suche. In den empirischen Erhebungen wurde dieser Zustand durch das *Self assessment*

---

<sup>12</sup> Siehe auch  $V_1$  – die Annahme der Existenz eines „nibs“ leitet sich erstens aus der Theorie zu den *casual-leisure* Szenarios nach [Elsweiler und Wilson, 2010] und zweitens als Äquivalent zu einem konkreten Informationsbedürfnis ab.

*Manekin* [Bradley und Lang, 1994] erfasst. Der Benutzer gibt in diesem Fall Selbstausskunft über seine aktuelle Stimmungslage. Zusätzlich wurden affektive Reaktionen mit einem EMG aufgezeichnet. Zu den Methoden siehe ab Kapitel 11.1.1.

**V<sub>4</sub>:** Variable vier erfasst in zwei Komponenten die Zufriedenheit des Nutzers mit der Suche:  $Z_1$  steht für die Zufriedenheit mit dem Ergebnis des Tasks (inklusive Erfüllung des Nicht-informationsbezogenen Suchziels).  $Z_2$  beschreibt die Zufriedenheit mit dem Suchverlauf.

Die beiden Untervariablen müssen nicht zwangsweise miteinander korrelieren, da theoretisch sowohl starke Ausprägungen von  $Z_1$  (Suchziel wurde erreicht, z.B. alle gewünschten Informationen gefunden) in Kombination mit einer schwachen Ausprägung von  $Z_2$  (z.B. es war umständlich die gewünschten Informationen zu finden, es dauerte zu lang, die gewünschten Informationen aus einer großen Menge an Inhaltselementen herauszufinden etc.) möglich sind.

**V<sub>5</sub>:** Ergänzt  $V_4$  um den quantitativen Aspekt, der angibt, in wie weit die Suche erfolgreich war hinsichtlich der gesuchten Information.  $E_{S1+S2}$  bezieht sich dabei rein auf die Menge der gefundenen Information im Verhältnis zum vorher definierten Informationsbedürfnis, wobei  $E_{S3}$  erfasst, in wie weit das Nicht-Informationsbezogene Suchziel erreicht wurde.

**V<sub>6</sub>:** Vorkenntnis: Bei der Vorkenntnis des Nutzers wird nach zwei Aspekten unterschieden: Zum einen, wie umfangreich seine Erfahrung im Umgang mit der Wikipedia sind und zweitens, ob er auf dem durchsuchten Themengebiet inhaltliche Vorkenntnisse hat.

Der jeweilige Wissenstand bzw. die Wissensstruktur des Nutzers sowie die Ergänzung der Wissensstruktur durch die präsentierten Informationen sind unsichtbare Variablen, die durch vorherige Abfragen (Selbsteinschätzung der Probanden im Laborexperiment) oder durch Schätzung auf Grund seines Verhaltens approximiert werden können [Marchionini, 1989; Kelly et al., 2002].

Da hier die Zufriedenheit mit der Informationspräsentation im Vordergrund steht, sind Erkenntnisse über die Wissensstruktur des Nutzers nur insofern relevant, als sie

erkennbar das Verhalten beeinflussen (z.B. ob ein Nutzer einleitende oder zusammenfassende Texte liest oder gleich zu Details im Artikel wechselt, sich offensichtlich im Themengebiet auskennt).

V<sub>L</sub>: Für den Lerntyp gelten die Definitionen nach [Felder, 1993].

V<sub>D</sub>: Demographische Faktoren sind Faktoren wie Geschlecht, Alter und höchster Bildungsabschluss.

V<sub>c</sub> : Komplexität des Suchtasks. Byström und Järvelin [Byström und Järvelin, 1995] haben in ihren Untersuchungen fünf Komplexitätsklassen identifizieren können, diese werden hier zur Bewertung der Aufgaben verwendet.

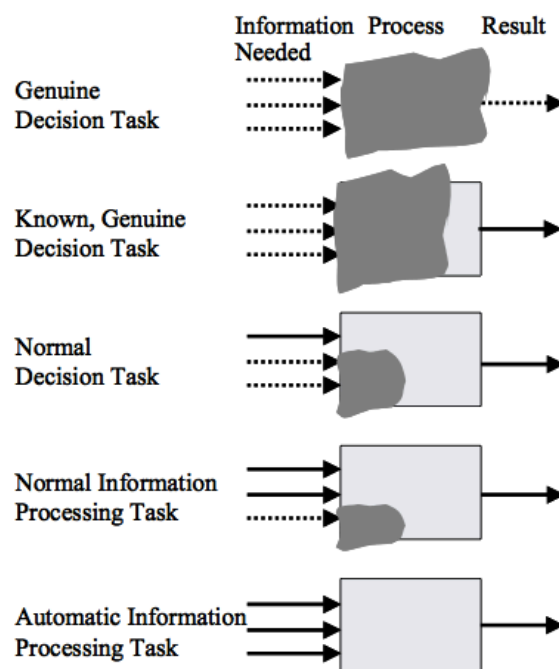


Abbildung 6-1 Task-Komplexitäts Klassifikation nach [Byström und Järvelin, 1995], Quelle: [Ingwersen und Järvelin, 2005, S.75]

### 6.2.2 Der weitere Nutzerkontext

Der zweite Baustein des Modells ist der taskbezogene weitere Kontext.

*Tabelle 6-2 Modell des weiteren Kontext*

#	Name	Wertebereich
V <sub>7</sub>	Domain	D <sub>t</sub> = {Klassifikation zur inhaltlichen Verortung der Artikel}
V <sub>8</sub>	verwendetes Gerät	De = { Laptop, PC, Smartphone, Tablet}
V <sub>9</sub>	verfügbare Zeit	T = {0...5}

#### ***Beschreibung der Variablen im Modell:***

V<sub>7</sub>: Die Variable Domain beschreibt den Wissensbereich, aus dem das Thema der Suche stammt. Dazu zählt nicht nur der inhaltliche Bereich, sondern auch der Gegenstand des Artikels, ob es sich um eine Person oder ein Gebäude, Ereignis oder ein anderes Lebewesen wie Pflanze oder Tier handelt. Um eine einheitliche Festlegung des Wertebereichs dieser Variable zu ermöglichen, wird in dieser und den damit verbundenen Arbeiten jeweils die in dieser Arbeit entwickelte Klassifikation (siehe ab Kapitel 9.7.3) zur Sortierung der Themen und Gegenstände verwendet.

V<sub>8</sub>: erfasst das zur Suche verwendete Gerät. Entscheidend sind hier Faktoren wie Displaygröße und Rechenleistung, aber auch Bedienung (z.B. mittels Maus und Tastatur oder Touchscreen).

V<sub>9</sub>: Verfügbare Zeit zur Bearbeitung eines Suchziels im Sinne von sehr wenig bis viel Zeit (0= sehr wenig, 5= viel oder unbegrenzt) da nach [Liu et al., 2012] die verfügbare Zeit einen Einfluss auf den Nutzer hat.

### 6.2.3 Taskdurchführung und Suchverhalten

Modellierung des Tasks und des Suchverhaltens:

Tabelle 6-3 Modell des Suchverhaltens

#	Name	Wertebereich
V <sub>10</sub>	Aktion	$\alpha_t = \{\text{read, scan, navigate, examine}\}$
V <sub>11</sub>	task completion	$C_{t_0} = \{0\%\}$ $C_t = \{0\% - 100\%\}$ gewünschte Information gefunden $\{0...5\}$ als Ersatzwert
V <sub>12</sub>	betrachtetes Element	$E_t = \{\text{text, pic, list, headline, introduction, infobox, Infografik, Index, link}\}$
V <sub>12B</sub>	betrachtetes Bild	$B_{1-n} = \{\text{Klassifikation für Bilder}\}$
V <sub>13</sub>	Relevanz [E]	$R_{E_t} = \{0...5\}$
V <sub>14</sub>	Suchphase	Entsprechend [Kuhlthau, 1991]
V <sub>15</sub>	Verbleibende Zeit	$T_{t_0} = 100\%$ $T_t = \{0\% - 100\%\}$

V<sub>10</sub>: Die Variable „Aktion“ bezieht sich auf die Aktionen, die der Nutzer bei der Suche durchführt. Diese werden in den Studien (siehe ab Kapitel 11.3.2) durch die Videolabels dargestellt und beziehen sich auf den jeweiligen Typ der Blickbewegungen [Duchowski, 2007] über die dargestellten Inhalte.

V<sub>11</sub>: *Task completion* beschreibt, in wie weit der Task zum Zeitpunkt  $t$  bereits erfüllt wurde. Diese Variable kann nicht direkt erfasst werden. Abhängig vom Aufgabentyp (bzw. der Endlichkeit und dem Grad der Definiertheit der gesuchten Information so wie den nicht-informationsbezogenen Zielen) ist die erreichbare *completion* eines Taks a priori schwer anzugeben, in den Studien gilt das subjektive Konzept, dass der Nutzer den Task beendet, wenn er selbst sich ausreichend informiert fühlt bzw. wenn die verfügbare Zeit abgelaufen ist.

Um Auskunft über die subjektiven Werte für V<sub>11</sub> zu bekommen, wird dies in den Untersuchungen nach jedem Task abgefragt.

V<sub>12</sub>: Die Variable „Betrachtetes Element“ bezieht sich auf das jeweils zum Zeitpunkt  $t$  im Fokus des Nutzers liegende Inhaltselement. Da das Set der Inhaltselemente in der

deutschsprachigen Wikipedia endlich ist, wurde dafür ein festes Inventar an Videolabels erstellt, die zu jedem Zeitpunkt das gerade vom Nutzer betrachtete Inhaltselement beschreiben und den Wertebereich der Variablen darstellen.

V<sub>12B</sub>: Beschreibt die betrachteten Bilder anhand einer Klassifikation für Bilder, aufgrund der situationsabhängigen Eignung von Bildtypen (siehe [Knäusl und Schubart, 2013; Schubart, 2012]).

V<sub>13</sub>: Die Relevanz eines Inhaltselements bezieht sich auf die subjekt- und situationsrelative Relevanz. Verschiedene Studien zeigen, dass dafür hauptsächlich die Verweildauer des Nutzers beim entsprechenden Element (bzw. der entsprechenden Seite) ausschlaggebend ist [Claypool et al., 2001; Buscher et al., 2008]]. In den hier durchgeführten Studien wird die Relevanz anhand der Häufigkeit des Auftretens der Variablenkombinationen V<sub>12</sub> und V<sub>10</sub> ermittelt.

V<sub>14</sub>: Die verbleibende Zeit für einen Task ist zum Zeitpunkt  $t_0$  100% und nimmt im Laufe der Suche ab. Allerdings muss V<sub>14</sub> nicht zwangsweise am Ende der Suche 0 % erreicht haben, denn ein Task kann bei Erfolg bereits früher abgebrochen werden. Andererseits kann die verbleibende Zeit bereits bei 0% sein, ohne dass sich ein (vollständiger) Sucherfolg eingestellt hat.



### Modell schematisch dargestellt

Die schematische Darstellung zeigt die Variablen zueinander und im Suchverlauf.

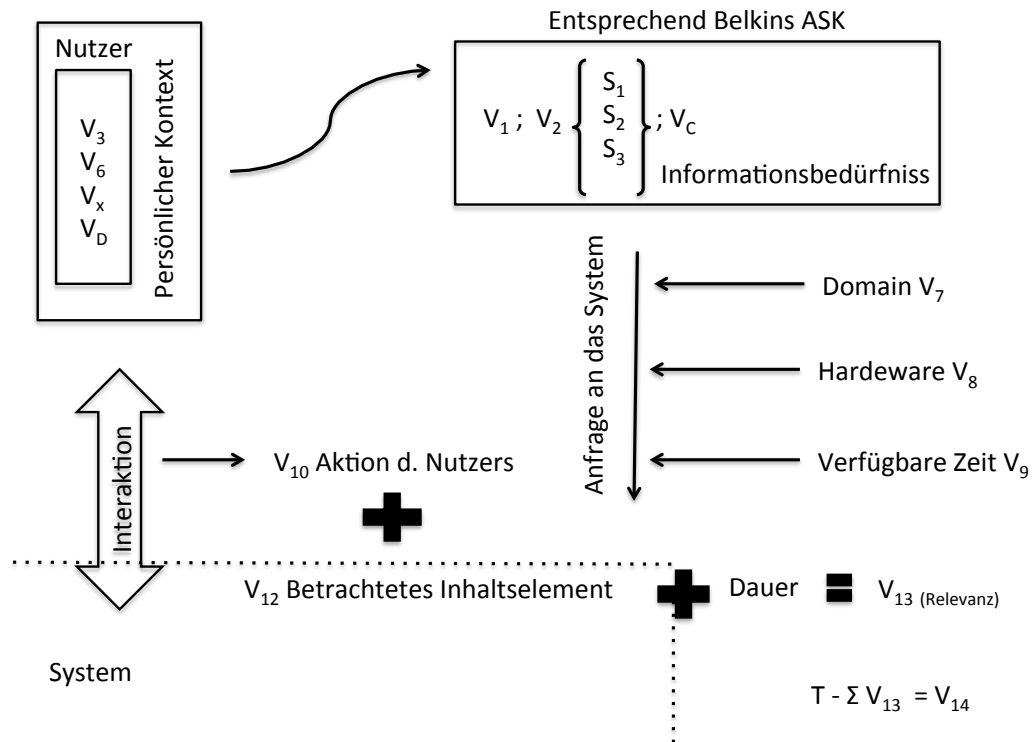


Abbildung 6-2 Schematische Darstellung des Kontextmodells

Der Benutzer befindet sich in seinem persönlichen Kontext, dazu gehören demographische Faktoren, seine emotionale Lage und seine jeweilige Vorkenntnis sowie die generellen Präferenzen entsprechend [Felder, 1993]. Die Interaktion mit den Inhalten des Systems ist Gegenstand der Untersuchungen. Sein Informationsbedürfnis wird durch die einzelnen Variablen definiert. Weitere Variablen, wie das verwendete Gerät, die durchsuchte Domain und die verfügbare Zeit, determinieren den weiteren Kontext der Informationssuche.



## 7 Berücksichtigung der Nutzungsszenarien

Die Existenz unterschiedlicher Nutzungsszenarien wurde hier dargelegt. Die daran anschließende Fragestellung beschäftigt sich nun damit, ob und wenn ja, warum es sinnvoll ist, diese bei der Informationspräsentation in Information Retrieval im Allgemeinen und bei der Wikipedia im Speziellen zu berücksichtigen.

Dazu werden die Theorien des *cognitive viewpoint* [Ingwersen und Järvelin, 2005] und der situativen Relevanz [Buckland, 1991; Ingwersen, 1982; Borlund und Ingwersen, 1997] besprochen, die eine grundlegende theoretische Fundierung der Berücksichtigung von Nutzersituationen aus kognitiver und informationstheoretischer Sicht bieten. Es wird weiter diskutiert, dass die Zufriedenheit der Nutzer als wichtigstes Ziel zu sehen [Ingwersen und Järvelin, 2005; Kelly und Fu, 2007; Belkin, 1980] und damit eine Berücksichtigung seiner individuellen Vorstellung der Zielerreichung im Information Retrieval plausibel ist.

Die *Verbesserung* der Mensch-Maschine-Interaktion durch adaptive Systeme ist eine Grundlage der Argumentation *für* eine Berücksichtigung der unterschiedlichen Nutzungsszenarien im Information Retrieval. Studien (ab Kapitel 7.2.1) zum adaptiven Information Retrieval, die vor allem auf semantischen Erweiterungen basieren, zeigen, dass Ergebnisse durch Berücksichtigung der Nutzerinteraktion bzw. dessen Suchhistorie und/oder der Verwendung erweiterter semantischer Konzepte verbessert werden konnten.

### 7.1 *Cognitive Viewpoint* und *Situational Relevance*

Grundsätzlich könnte man den Ansatz der situativen Relevanz und der subjektiven Nutzerbewertungen durchaus in Frage stellen. Denn der Abgleich von Termen nach den Methoden des Information Retrieval<sup>13</sup> überprüft, ob ein Term aus der Query des Nutzers in einem Dokument vorhanden ist, bzw. in einem entsprechendem Verhältnis zur Gesamtzahl der Terme auftritt. So kann eine inhaltliche Relevanz berechnet wer-

---

<sup>13</sup> Zum Beispiel mengentheoretisch, algebraisch oder probabilistisch, siehe dazu [Lewandowski, 2005].

den (z.B. mit Hilfe von Textstatistik, dem Booleschen Modell (inkl. Gewichtung) oder linktopologische Verfahren). Durch bestimmte Maßzahlen im Information Retrieval (bezüglich Relevanz der gefundenen Dokumente und Vollständigkeit der Treffermenge) kann die Güte der Treffermenge präzise berechnet und ein inhaltliches Ranking vorgenommen werden [Lewandowski, 2005]. Dennoch gibt es plausible Theorien dafür, dass diese Güte der Treffermenge nicht allein von der Precision und dem Recall abhängt. Dabei spielt der sog. *cognitive viewpoint* bzw. der Aspekt der situativen Relevanz und der Information als subjektives, adressatenbezogenes Konzept eine wichtige Rolle.

### 7.1.1 Information als subjektives Konzept

Wie Gernot Wersig [Wersig, 1971] feststellt, gibt es annähernd so viele Definitionen des Informationsbegriffs, wie es Autoren gibt, die darüber schreiben. Auch ohne genauer auf die Typologisierung des Informationsbegriffs einzugehen oder eine terminologische Grundsatzdiskussion darüber zu führen, lässt sich feststellen, dass sich dieser auf unterschiedliche Objekte beziehen kann (z.B. Information als impliziter Teil der Materie auf der Erde, also schon vorhanden, Information als Teil eines „Wissenvorrats“ des Senders, Information als Sendung, also physikalisches Ereignis und/oder Konglomerat an Zeichen, Information als Bedeutung einer Sendung oder Information als Veränderung des Wissens des Empfängers [Wersig, 1971, S. 30-34]). Die verschiedenen Definitionen von Information unterscheiden sich teilweise in einem hier relevanten Punkt: Information als subjektives und damit erst im Kontext des Empfängers relevantem oder Information als objektives Konzept.

Die naturalistische Sichtweise von Information (hier vor allem [Hartley, 1928]) bezieht sich auf die Nachrichtenübermittlung zwischen Maschinen. Hier gibt es explizit keinen pragmatischen und handlungsrelevanten Aspekt von Information. Diese Definition von Hartley bezieht sich nur auf einen physikalischen Vermittlungsprozess [Capurro, 2000]. Innerhalb der Informationswissenschaft ist diese Interpretation des Informationsbegriffs auf Grund der Tatsache, dass die *Bedeutung* der Information dabei keine Rolle spielt nicht geeignet. Bei der Interaktion von Menschen mit Information steht der pragmatische Aspekt an vorderster Stelle, wie im Folgenden kurz dargestellt wird.

Hierzu sei zunächst die Sichtweise von Buckland [Buckland, 1991] erwähnt, der argumentiert, dass nicht die *angebliche* Wahrheit einer Information entscheidend ist sondern

das, was der Rezipient für wahr hält. Informationsdinge sind also dann informativ, wenn sie in einer bestimmten Situation vorliegen, was den *potentiell* informativen Charakter von Informationseinheiten betont. Hier ist allerdings einzuwenden, dass zwar zutrifft, dass nicht festzustellen ist, ob eine Information *an sich* wahr ist [Capurro, 2000], dass aber grundlegend Informationen immer mit wahren, höherrangigen *beliefs* [Herrschaft, 1996, S. 175, zitiert nach Capurro, 2000] stimmig sein müssen, damit wir sie für wahr halten. Damit kann nicht ausschließlich von einer rein *subjektiven* Wahrheit der Information ausgegangen werden. Für diese Arbeit gilt allerdings diese Wahrheit, die sich darauf gründet, dass der Rezipient die Information für wahr hält [Buckland, 1991], da hier im Information Retrieval Prozess in der Wikipedia nicht überprüft wird, ob die gefundene Information plausibel ist, also nicht im Widerspruch zu anderen, als wahr bestätigten Fakten steht. Dies dient dazu, die Komplexität der Fragestellung zu reduzieren und basiert auf der Annahme einer hinreichenden Qualität der Inhalte, die die Richtigkeit mit einschließt, siehe dazu auch Kapitel 9.6.

Damit gilt in diesem Punkt die Sichtweise von Buckland [Buckland, 1991], allerdings lässt dieser eine konkrete Definition von Information vermissen. In dieser Lücke kann man Seifert [Seifert, 1971, S. 76] zitieren, der konkret vorschlägt:

*„Information ist hiernach eine gegenwarts- und praxisbezogene Mitteilung über Dinge, die mir im Augenblick zu wissen wichtig sind“*

Damit ist ein entscheidender Aspekt von Information angesprochen: Der Praxisbezug, der entsprechend die Relevanz oder überhaupt die Verwendbarkeit der Information für den Nutzer beeinflusst, also eine pragmatische Sichtweise. Die Definition sagt weiter, dass es um *„Dinge, die mir im Augenblick zu wissen wichtig sind“* [Seifert, 1971, S.76] – womit die Situationsrelativität betont wird.

Ergänzend lässt sich weiter die Sichtweise von Rainer Kuhlen [Kuhlen, 2004, S. 14] anwenden:

*„Aus informationswissenschaftlicher Sicht einigt man sich halbwegs auf eine pragmatische Sichtweise: Bedeutung, Handlungsrelevanz und damit Nutzen von Information sind entscheidend“*

Auch nach Kuhlens [Kuhlen, 2004] Interpretation wird die Information erst in einem Nutzungskontext aktiv: Sie ist kontextabhängig und adressatenbezogen.

Den genannten Definitionen entsprechend ist im Fall dieser Arbeit alles als informativ bzw. als Information zu betrachten, das dem Suchende dazu dient, seinen ASK [Belkin, 1980] zu beseitigen. Im Umkehrschluss bedeutet das, dass nach dieser Definition diejenigen Daten, die diesen Nutzen situativ nicht haben und damit auch nicht handlungsrelevant sind, keine Information für den Nutzer darstellen. Das zu betonen ist wichtig, da die Unterscheidung zwischen für den Nutzer informativen Inhaltselementen und nicht-informativen Angeboten für eine Untersuchung der situativen Informationsrezeption entscheidend ist.

Die in der Wikipedia abgespeicherte Menge an Fakten (ob wahr oder falsch) bleibt davon als potentielle Information natürlich unberührt. Die Unterscheidung beginnt erst, wenn aus einer Teilmenge des in der Wikipedia in Form von Daten gespeicherten Wissens Information wird/werden soll, also in einer konkreten Situation angefordert wird. Eine nicht von der Definition von Information ausgehende sondern von der anderen Seite, dem Menschen, aus argumentierende aber entsprechende Theorie ist die des sog. „cognitive viewpoints“ [Ingwersen und Järvelin, 2005]

### **7.1.2 Cognitive Viewpoint**

Der *cognitive viewpoint* ist eine alternative Sichtweise zu der „klassischen“ Haltung im Information Retrieval, die sich vor allem auf Systemebene und Dokumente konzentriert. Im Prinzip basiert die kognitive Sichtweise darauf, dass bei der Verarbeitung und Bewertung von Information vor allem auch die jeweilige menschliche Wahrnehmung und das subjektive Verstehen eine Rolle spielen – nicht nur die binäre Sicht von Relevanz [Borlund, 2000]. Bei dieser Herangehensweise werden auch das vorhandene Wissen und die kognitive Struktur des Menschen bei der Bewertung von Inhalten und

der Vorgehensweise bei der Informationssuche berücksichtigt. Für diese Arbeit ist der *cognitive viewpoint* eine theoretische Basis, da zusätzlich zur systemseitigen Sicht von Relevanz, die auf den probaten Verfahren des Information Retrievals aufbaut, auf die individuellen Nutzervorstellungen referiert wird.

Vor allem die Berücksichtigung der *Wechselwirkung* zwischen Information und Nutzer ist hier wichtig [Belkin, 1977]. Diese Wechselwirkung führt dazu, dass es keine absolute Relevanz einer Dokumenteinheit gibt, sondern diese sich dynamisch während des Suchprozesses verändert und darüber hinaus durch äußere Faktoren beeinflusst wird. Daraus ergibt sich wieder die Frage, wie die Information durch den einzelnen User in Abhängigkeit des aktuellen Aufgaben-Kontextes (inklusive des konkreten Suchziels) bewertet wird. Belkin ist hier später noch etwas genauer [Belkin, 1980] und arbeitet mit dem *anomalous state of knowledge* [ASK], im Prinzip also einer Vorstellung des mentalen Modells eines Nutzers inklusive dessen Vorwissen und seinen konzeptuellen Vorstellungen der Welt, die in einem Ausnahmezustand sind, da eine Lücke in der Wissensstruktur besteht. Diese Lücke in der Wissensstruktur unter Berücksichtigung der weiteren Aspekte seiner Situation entspricht dem konkreten Informationsbedarf.

Auch Carol Kuhlthau spricht diesen Punkt an [Kuhlthau, 1991, S.1]:

*„Information systems and the intermediaries who manage them traditionally have been driven by a bibliographic paradigm centered on collecting and classifying texts and devising search strategies for their retrieval. This has promoted a view of information use from the system’s perspective and has concentrated retrieval on questions that best match the system’s representation of text rather than responding to user’s problems. The bibliographic paradigm is based on certainty and order, whereas users’ problems are characterized by uncertainty and confusion“*

Kuhlthau betont damit die Lücke, die zwischen der vom System angenommenen und bereitgestellten Wissensstruktur und derjenigen Information, die der Nutzer tatsächlich benötigt, um sein persönliches Informationsbedürfnis zu stillen bzw. seine eigene vorhandene Wissensstruktur zu erweitern, besteht. Sie stellt die formal organisierte

Wissensstruktur eines Informationssystems der viel weniger formal organisierten Abbildung des Wissens (und Formulierung der Wissenslücke) auf der Seite des Nutzers gegenüber [Kuhlthau, 1991]. Daraus resultiert die Folgerung, die Repräsentation der Ergebnisse einer Suche ist „...coming from the system's world...“ [Kuhlthau, 1991, S. 1] und damit nicht der Vorstellung und Welt des Nutzers entsprechend. Damit gilt auch hier die binäre, systemseitige Relevanz von Informationen nur mehr eingeschränkt. Die Sichtweisen und das Verständnis des Nutzers spielen ebenfalls eine zentrale Rolle.

Ingwersen und Järvelin zeigen, dass der sog. *cognitive viewpoint* gut und unter Umständen auch besser als eine philosophische oder anthropologische Sichtweise geeignet ist, um die spezielle Problematik der Subjektivität von Relevanz beim IR erstens zu zeigen und zweitens auch ein Stück weit zu lösen [Ingwersen und Järvelin, 2005].

Diese holistische Sichtweise ersetzt nicht die pragmatischen und rationalen Ansätze, sondern *ergänzt* sie [Ingwersen und Järvelin, 2005]. Das Konzept des *cognitive viewpoint* soll vor allem dabei helfen, Faktoren wie subjektive Relevanz und Rezeption oder Entwicklung des Suchbedürfnisses über die Zeit besser zu beschreiben und in den Retrieval-Prozess mit *einfließen* zu lassen.

[Ingwersen und Järvelin, 2005, S. 25] beschreiben fünf wesentliche Dimensionen des *cognitive viewpoint*:

- *Information processing takes place in senders and recipients of messages;*
- *Processing takes place at different levels;*
- *During communication of information any actor is influenced by its past and present experiences (time) and its social, organizational and cultural environment;*
- *Individual actors influence the environment or domain;*
- *Information is situational and contextual.* (Hervorhebungen der Autoren)

Die Autoren [Ingwersen und Järvelin, 2005] betonen hier explizit die Situations- und Kontext-Bezogenheit von Information, entsprechend [Kuhlen, 2004; Buckland, 1991; Seifert, 1971]. Der zweite wesentliche Aspekt, den Ingwersen und Järvelin für die kognitive Sichtweise nennen, ist der Einfluss des Kontextes auf die Akteure – die Kommunikation von Information ist immer unter Berücksichtigung dieses Einflusses zu sehen.



In dieser Arbeit entspricht dem die Hypothese, dass das situationsbezogene Nutzerverhalten bei der Suche in der Wikipedia unterschiedlich ist.

Wersig [Wersig, 1971] konstatiert weiterhin, dass das Konzept des allgemeinen Informationsbedürfnisses immer der individuellen Sichtweise des Einzelnen in Kombination mit seiner jeweiligen Situation unterliegt. Jedes Individuum hat jeweils eine eigene Wahrnehmung seiner „Problemsituation“ (vgl. auch [Belkin, 1980]) und davon, wie er die daraus resultierende Unsicherheit verringern will.

Der geforderten Berücksichtigung der individuellen kognitiven Vorstellungen der Einzelperson und die Situationsbezogenheit der Informationssuche und -rezeption folgend ist eine adaptive Informationspräsentation im Information Retrieval plausibel.

### 7.1.3 Situative Relevanz

Die Überlegungen dieser Arbeit zeigen, dass Relevanz – obwohl eines der wichtigsten Kriterien zur Evaluierung von Ergebnissen eines Retrieval Prozesses – von Seiten des Nutzers kein einheitlich klar definiertes und vor allem nicht universell anwendbares Konzept ist.

In den Studien dieser Arbeit bedeutet Relevanz, dass der Nutzer subjektiv das jeweilige Inhaltselement für seine konkrete und individuelle Aufgabe in Kombination mit seinen persönlichen Präferenzen als relevant betrachtet. Zur Begründung sei hier wieder auf [Belkin, 1980, S. 134] verwiesen:

*„c) the user examines the text(s) presented by the mechanism and the need is satisfied completely, partially, or not at all. The user's judgement as [...] the contribution of each text in satisfying the need establishes the usefulness or relevance of that text to the need“*

Bei der Suche in der Wikipedia bedeutet Relevanz also, dass der Nutzer das jeweilige Inhaltselement in die Lösung seiner Aufgabe mit einbezogen hat, indem er ihm Aufmerksamkeit schenkte.

[Buckland, 1991] schreibt zu diesem Punkt, dass Informationsdinge zwar durchaus eine generell bestimmbare *potentielle* Informativität haben, die in einem allgemeineren Konsens festgestellt werden kann. Aber die potentielle Informativität ist eben nur eine Basis – tatsächlich informativ sind diese erst, wenn sie in einer bestimmten Situation angetroffen werden. Dies begründet er wiederum darin, dass *Information-als-Prozesse* situationsgebunden ist [Buckland, 1991]. Somit stehen sich zwei zunächst getrennte Konzepte gegenüber, die erst durch die Situation verbunden werden: ein potentiell informatives Informationsding und Information-als-Prozess. Sprich: das Informationsding ist vorhanden, und wenn es dann in einer bestimmten Situation angetroffen wird, bekommt es einen prozessual-informativen Charakter. Als Situation wird in dieser Arbeit wiederum der Nutzungskontext, bestehend aus den in Kapitel 1 zusammengefassten Faktoren verstanden.

Auch wenn die potentielle Informativität von Informationsdingen (in der vorliegenden Arbeit z.B. die Inhaltselemente) generell situationsunabhängig und durch kulturelle und intellektuelle Übereinkunft ein Stück weit festgelegt ist, so ist doch die Situation, in der sie angetroffen werden, um bei der Terminologie Bucklands [Buckland, 1991] zu bleiben, entscheidend über ihre Relevanz bzw. tatsächliche Informativität.

Nach [Wang und Soergel, 1998] ist weiter auch der Neuigkeitswert einer Information und die Passgenauigkeit in das mentale Modell des Nutzers [Ingwersen und Järvelin, 2005] entscheidend für die Relevanz einer Information.

Kuhltau [Kuhltau1991] zitiert zur Lösung dieses Problems [James, 1983; Hall, 1981 und Ingwersen, 1982] mit dem gemeinsamen Ansatz, deshalb erstens die Integration der Ergebnisse in die Wissensstruktur des Nutzers und zweitens dessen eigene Bewertung der Nützlichkeit der Information zu berücksichtigen.

Auch Linda Schamber und ihre Kollegen [Schamber et al.,1990] betonen den Sinn eines dynamischen, nutzerzentrierten und situativen Ansatzes. Sie sehen im Konzept der Relevanz mehr als eine Dimension zu berücksichtigen: nämlich einerseits die „*internal factors*“ – was in etwa dem kognitiven Standpunkt (nach [Ingwersen und Järvelin,

2005]) entspricht und andererseits die „*external factors*“ – also den situationsgebundenen Fakten [Schamber et al., 1990].

Daraus resultiert zusammenfassend, dass die Relevanz eines Informationsdings [Buckland, 1991] aus dem situationsgebundenen Urteil des Nutzers und unter Berücksichtigung seiner bereits vorhandenen Wissensstruktur entsteht.

## 7.2 Zufriedenheit des Nutzers als oberste Priorität

Auch eine nicht adaptive Informationspräsentation in der Wikipedia ist bei korrekten Suchtermen und validem Suchalgorithmus im System sachlich richtig. Die genannten Schwierigkeiten bei der individuellen Selektion der Inhalte durch den Nutzer ändern daran nichts. Allerdings ist die *Nutzerzufriedenheit* nicht unbedingt gesichert. In der Forschung zum Information Retrieval gibt es Argumentationen, die explizit dafür sprechen, die subjektive, kontextabhängige Zufriedenheit des Nutzers (zusätzlich zur sachlichen Richtigkeit der Ergebnisse) als wichtigstes Indiz für eine erfolgreiche Suche zu betrachten (siehe u.a. bei [Ingwersen und Järvelin, 2005; Lau et al., 2008; Kelly und Fu, 2007; White und Kelly, 2006]). Auch Huffman und Hochster [Huffman und Hochster, 2007] zeigen mit einer Studie, dass die Zufriedenheit der Nutzer nicht nur allein von der entsprechenden Relevanz der gefundenen Ergebnisse abhängt, der Verlauf der Suche spielt dabei ebenfalls eine wichtige Rolle.

[Ingwersen und Järvelin, 2005] plädieren in ihrem Buch „*The turn: Integration of information seeking and retrieval in context*“ für eine stärkere Berücksichtigung des Retrieval Kontextes bzw. sogar für eine situative, zu jeder Suchanfrage neu bewertete Kontextsituation.

Denn nach [Blair, 1990] wird die Qualität einer Information nicht (nur) nach ihrer *Korrektheit* beurteilt, sondern vor allem nach ihrer *Adäquatheit* beurteilt.

Auch im Zusammenhang mit der vorliegenden Arbeit, die die Nutzung der Wikipedia (bzw. der deutschsprachigen Wikipedia) untersucht, erscheint das Plädoyer gerechtfertigt. Ein Stück weit wird dies bereits durch bestimmte Strukturelemente der Wikipedia Artikel zu erreichen versucht, die wahlweise Übersichten oder ausführliche Informati-

onen bieten. Umfang und Länge der Artikel allerdings sind statisch und damit nicht kontextabhängig präsentiert. Das widerspricht der Forderung von [Belkin, 1980, S. 134]:

*„This focus on whole text or combinations of text also arises from my view of the function of an IR System: to help people to solve problems, rather than directly to solve problems posed to them [...]“<sup>14</sup>*

Im Prinzip versucht jede Art von IR System zunächst, die vom Nutzer formulierte Suchanfrage bestmöglich zu beantworten und höchste Nutzerzufriedenheit zu erreichen. Das „Problem“ besteht darin, dass dem System erstens wesentliche Informationen über den Nutzer und seine Anfrage fehlen [White und Kelly, 2006].

So auch [Swanson, 2004, S. 261]:

*„Relevance is a user judgement that goes beyond the mere measure of presence, frequency, or absence of a term in a document. It is rather a factor of what concepts the terms represent, the relations between these concepts, and how users interpret them , and how they relate to the information in the document“*

In diesem Punkt ist sich die aktuelle Forschungslage einig. Der Nutzer hat sein ganz eigenes „Problem“. In der Kommunikation mit dem System gehen Informationen verloren bzw. erreichen das System nur auf eine ihm nicht verständliche Art und Weise. Der Nutzer urteilt subjektiv über Nutzen und Relevanz in *Abhängigkeit* seiner Situation [Swanson, 2004]. Um die Zufriedenheit des Nutzers über die sachliche Richtigkeit der Informationen hinaus zu gewähren, muss ebenfalls die Adäquatheit der Information gewährleistet sein [Blair, 1990].

---

<sup>14</sup> Belkin bezieht sich hier auf den „*Anomalous state of knowledge*“ [Belkin, 1980] und die Relativität der Relevanz von Dokumenten bzw. eben Dokumentteilen [in diesem Fall „*parts of text*“]. Betrachtet man die hier immer wieder auftauchenden sog. *casual leisure* szenarios, in denen der Nutzer zur Freizeit-Zwecken im Informationssystem surft formal ebenfalls als „Problem“ dann ist das Zitat umso treffender.

Wie [Ingwersen und Järvelin, 2005; Swanson, 2004; Blair, 1990 und Belkin, 1980] betonen, ist deshalb die der Situation angemessene Information wichtig, da Adäquatheit nicht universell ist.

[Marchionini, 1995] generalisiert mit seinem *human centred framework* diesen Aspekt, in dem er den Nutzer, dessen individuelle Interaktion und subjektive Bewertung der Ergebnisse, seine Motivation, die durchsuchte Domain, aber auch physische und kognitive Aspekte des Suchprozesses in den Mittelpunkt stellt. Dabei bezieht er sich hauptsächlich auf die Entscheidungen, die der Nutzer während der Suche trifft. *Human centred design* ist ein zentrales Konzept, das sich genau mit dieser Frage beschäftigt: Wie kann man den Nutzer darin unterstützen, seinen – durch äußere Faktoren beeinflussten – Bedürfnissen gerecht zu werden bzw. eben auch, wie kann ein System mit möglichst viel Wissen über diese Bedürfnisse und den Nutzer optimal designed werden. [Marchionini et al., 2000] beschreibt als einen wesentlichen Punkt von *human centred* in der Frage des Information Retrieval die Berücksichtigung der *real life tasks*, sprich, der Situationen, wie sie eben im „echten Leben“ und nicht nur aus Systemsicht, vorkommen.

[Marchionini et al., 2000, S.1]:

*„People are interested in accomplishing those tasks rather than executing searches“*

Die Thematik der *real life tasks* wurde hier bereits besprochen (siehe Kapitel 5). Die Forderung nach einer Berücksichtigung dieser zur Sicherstellung der Zufriedenheit der Nutzer ist nachvollziehbar [Marchionini et al., 2000].

Im Bereich der Mensch-Maschine-Schnittstelle gibt es Methoden, die helfen, den Nutzer in seinem *real live task* zu unterstützen [Benyon et al., 1994] und eine höhere Zufriedenheit mit den Ergebnissen zu erreichen. Diese beziehen sich allerdings hauptsächlich auf Funktionalitäten und Interfaces. So können etwa *advanced searches* angeboten, semantische Verknüpfungen zu ähnlichen Themengebieten [Berger et al., 2004] und/oder Visualisierungen von Suchergebnissen vorgenommen werden. Auch die adaptive Gestaltung der Inhalte bietet Chancen, die Zufriedenheit der Nutzer zu verbessern.

### 7.2.1 Adaptive Mensch-Maschine-Interaktion

Adaption ist in verschiedenen Bereichen der Mensch-Maschine-Interaktion (Funktionalität, Darstellung, Geschwindigkeit etc.) in verschiedenen Stufen der Anpassung möglich [Benyon et al., 1994].

Im eher einfacheren Fall werden „nur“ simple Systemeinstellungen (z.B. oft aufgerufene Funktionen, Navigationssystem im Auto regelt Helligkeit des Displays entsprechend den Lichtverhältnissen in der Umgebung) an die Gewohnheiten des Nutzers angepasst. Sensiblere und höher trainierte Systeme können aber beispielsweise Absichten, persönliches Vorwissen und Präferenzen des Nutzers berücksichtigen und unter Umständen auch je nach Domain, in der der Nutzer sich bewegt, reagieren [Benyon et al., 1994].

Für ein gutes adaptives Interface ist ausschlaggebend, dass erstens nur wirklich relevante Faktoren berücksichtigt werden und dass die Adaption dann auch tatsächlich die Effizienz und Effektivität des Systems – und im besten Fall zusätzlich den sog. *Joy of use*<sup>15</sup> – steigert. Deshalb wird die Hypothese, dass eine adaptive Informationspräsentation der Inhalte der Wikipedia sinnvoll ist, nur unter der Bedingung angenommen, dass erstens Adaption im Information Retrieval generell die Nutzerzufriedenheit steigert und zweitens unterschiedliche Präferenzen der Nutzer nachweisbar sind.

Wie und was angepasst wird, ist unterschiedlich. [Benyon et al., 1994] beschreibt eine Liste verwendbarer Techniken von der Nutzer-kontrollierten Adaption bis zur statistischen Analyse des Nutzerverhaltens. Im Unterschied zu adaptierbaren Benutzerschnittstellen, die der Nutzer nach seinen persönlichen Präferenzen aktiv gestalten kann, passen sich adaptive Systeme zur Laufzeit automatisch an den Nutzer an. Weiterhin kann man unterscheiden, ob die Präsentation der Inhalte oder die Funktionalitäten jeweils optimiert werden.

Trumbley et al. [Trumbley et al., 1994] zeigten in einer Studie eine deutliche Leistungssteigerung der Probanden durch Verwendung einer Software, die sich dem jeweiligen

---

<sup>15</sup> Spass an der Benutzung eines Systems, im Vordergrund stehen hier vor allem auch Ästhetik und Emotionen, vgl. [Hassenzahl, 2004].

Erfahrungsstand des Nutzers anpassen kann. Das ist ein deutlicher Hinweis darauf, dass Adaption die Zufriedenheit und Performance der Nutzer verbessern kann.

Auch beim E-Learning kann eine adaptive Informationspräsentation das Lerntempo des Nutzers berücksichtigen – und gleichzeitig sicherstellen, dass gewisse Lernziele erreicht werden (z.B. durch Wiederholen)[Kostadinova et al., 2012].

Adaptive Mensch-Maschine-Interaktion ist unter der Prämisse sinnvoll, dass ein Nutzer ein bestimmtes Ziel hat, das er mit Hilfe eines Systems erreichen will. Davon ist bei der Informationssuche in der Wikipedia auszugehen<sup>16</sup> (siehe dazu ab Kapitel 5.3). Unter Umständen gelingt ihm die Zielerreichung zunächst nicht (z.B. durch Selektionsprobleme [Morita und Shinoda, 1994; Wu und Wu, 2011]), dann ist es hilfreich, das System ist mit Möglichkeiten ausgestattet, um dem Nutzer helfen zu können. Dazu muss es das Ziel des Nutzers kennen und wissen, wie es reagieren kann, um ihn bei der Zielerreichung zu unterstützen. Dabei spielt es zunächst eine untergeordnete Rolle, um welche Art Informationssystem es sich dabei handelt: sowohl im E-Learning, als auch bei Informationssystemen wie Stadtführern auf dem Smartphone, Systemen zur Unterstützung des Alltags (z.B. Gesundheitsratgeber) oder Wissensdatenbanken, wie der Wikipedia, ist die Hilfe zur Zielerreichung (bzw. der besseren und einfacheren Zielerreichung) sinnvoll.

### 7.2.2 Adaption im Information Retrieval

Die Idee, die Inhalte bzw. Auswahl und Präsentation der vorhandenen Inhalte adaptiv und dem Kontext entsprechend zu gestalten, ist nicht neu. Durch die unüberschaubare Menge an Informationen motiviert gibt es bereits einige Ansätze, das Information Retrieval durch adaptives Verhalten eines Systems zu verbessern [z.B. Hu et al., 2009; Somol et al., 1999; Morita und Shinoda, 1994].

Im Wesentlichen kann im Information Retrieval die Anpassung auf drei Ebenen stattfinden: erstens bei der Formulierung der Query (das Problem des unscharfen Informationsbedürfnisses und des Informationsverlustes durch die Verbalisierung desselben berücksichtigend), zweitens auf der Ebene des Dokumentenabgleichs (z.B. durch semantische Erweiterungen oder Disambiguierung von Anfragetermen durch Berück-

---

<sup>16</sup> Sowohl inhaltlich als auch nicht-informationsbezogen.

sichtigung des semantischen Kontextes) oder drittens bei der Präsentation der Ergebnisse (wobei hier bei der aktuellen Forschungslage der Schwerpunkt tendenziell mehr auf der Gestaltung von Ergebnislisten im Sinne des Dokumenten-Rankings liegt, denn auf der Präsentation der Inhalte selbst).

Ein konkreter Ansatz zur Adaptivität im Information Retrieval ist, dass das System sich vorangegangene Anfragen merkt und so semantische Disambiguierung von Anfragetermen unterstützen kann. Lau, Bruza und Song [Lau et al., 2008] verwenden hierzu eine logikbasierte *believe-revision*. Das bedeutet, dass das „Umfeld“ des Nutzers im Sinne seiner bisherigen Aktivitäten, seinem Wissen (sofern bekannt) und der durchsuchten Domain unter der Anwendung von Logik in die Interpretation der Suchterme mit einbezogen wird. So erreichen die Autoren bessere Ergebnisse als mit vergleichbaren, nicht adaptiven Vorgehensweisen.

Abbildung 7-1 zeigt einen Überblick über den Aufbau eines adaptiven Information Retrieval Systems. Zwischen der Anfrage des Nutzers und den Informationsobjekten steht außer einem *matching mechanism* noch ein *learning mechanism*, der auf Basis eines Nutzer-Profiles und dem Feedback die Retrieval-Situation berücksichtigt, damit den *matching mechanism* dementsprechend beeinflusst und – wie sich zeigte – verbessert [Lau et al., 2008].

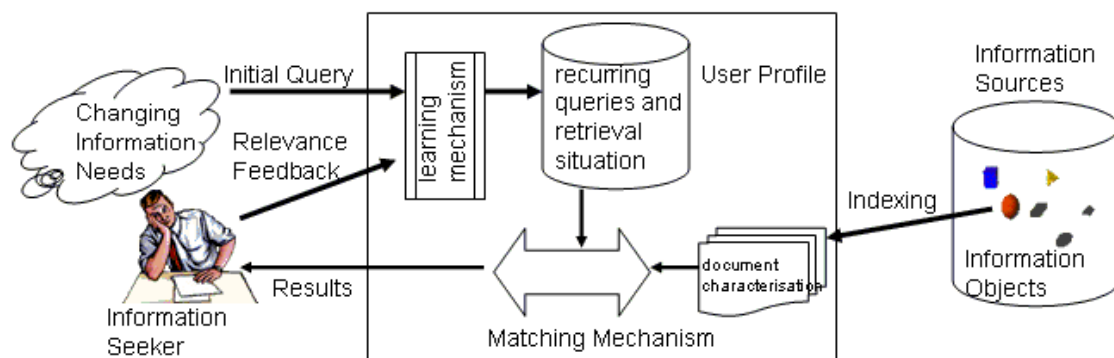


Abbildung 7-1 Überblick über ein adaptive gestaltetes Information Retrieval System [Lau et al., 2008, S.4]



Crestiani und van Rijsbergen [Crestani und van Rijsbergen, 1997] entwickelten auf der Basis logischer und konzeptueller Beziehungen ein Modell, das sie im Rahmen eines adaptiven Information Retrieval Systems implementierten. Sie testeten verschiedene Lernalgorithmen zur Verbesserung der Retrieval Ergebnisse und erzielten damit eine deutlich bessere Bewertung der Treffermenge durch die Nutzer. Die Adaption findet hier ebenfalls vor allem in einer Anpassung bzw. Erweiterung der Query und damit einem genaueren und adaptiven *matching* mit der Dokumentenmenge statt.

Shen [Shen, 2007] stellt in den Mittelpunkt ihrer Forschung die Frage nach einem aktualisierbaren Nutzermodell, das notwendige Informationen über den Kontext des Nutzers liefern soll, um eine Personalisierung der Suche zu ermöglichen. In ihrem Framework verwendet sie die Interaktion des Nutzers mit dem System im Sinne von *Implicit Relevance Feedback* (hier vor allem *clickthrough*), um sofort auf Veränderungen des Nutzerkontexts reagieren zu können und die Suche anzupassen. Dabei kommen vor allem statistische Sprachmodelle zum Einsatz, anhand derer zum Beispiel die bereits betrachteten Dokumente oder vorangegangene Anfragen mit in die Suche einbezogen werden können (*query expansion*). So kann das Ranking der Antwort-Dokumente entsprechend überarbeitet werden. Es gelang Shen, eine Client-seitige Anwendung zu implementieren, die durch eine Abspeicherung des bisherigen (semantischen) Nutzerkontexts auf Basis vorangegangener *queries* und verwendeter Dokumente jeweils ein re-ranking der von einer Web-Suchmaschine gelieferten Dokumente durchführt. Bei einer Vergleichsstudie ließ sich feststellen, dass die überarbeitete Reihenfolge der Ergebnisse Verbesserungen gegenüber dem ursprünglichen Ranking von Google zeigte.

Einen ähnlichen Ansatz verfolgen auch Wilrich et al. [Wilrich et al., 2006], die speziell im Bereich der digitalen Bibliotheken untersuchten, in wie weit die Erstellung eines Nutzerprofils, das vor allem die Metadaten der bereits besuchten Dokumente berücksichtigt, die Suchergebnisse für den Nutzer personalisiert und verbessert. Ein großer Vorteil hierbei ist, dass die Dokumente in digitalen Bibliotheken in der Regel mit sehr guten Metadatensätzen ausgestattet sind. Die umfangreichen Informationen über die Inhalte erleichtern eine Präzisierung der Auswahl. Diese Arbeiten machen deutlich, dass Adaption mit ausreichend Wissen über den Nutzer, hier vor allem über den semantischen Kontext seiner Anfrage, großes Potential hat, die Nutzerzufriedenheit zu verbessern.

Berger, Dittenbach und Merkl [Berger et al., 2004] zeigen, wie sie mit Hilfe assoziativer Verknüpfungen in Form eines Wissensrepräsentationsmodells semantische Relationen als Grundlage für ein mehrsprachiges Information Retrieval System verwenden. Adaption findet hier auf zwei Ebenen statt: erstens in Form eines *matching* mit den Suchtermen, das durch (gewichtete) semantische Relationen mit anderen Termen verbessert wird und dadurch zweitens eine sog. *query-expansion* (durch Einbeziehen der erweiterten semantischen Beziehungen) bewirkt. [Berger et al., 2004] versuchen so vor allem, einen eher intuitiven Zugang zu den Informationen zu schaffen.

Wen, Zhou und Aggarwal [Wen et al., 2007] wendeten eine adaptive und adaptierbare Methode an, indem sie auf der einen Seite dem Nutzer die Möglichkeit gaben, selbst aktiv einen semantischen Suchkontext mitzugestalten und zweitens das System diesen Prozesse unterstützte. Auch diese Autoren konnten zeigen, dass die Maßnahmen die Relevanz der erhaltenen Ergebnisse für den Nutzer tatsächlich verbesserten.

Diane Kelly und Xin Fu [Kelly und Fu, 2007] präsentieren, dass mit Hilfe von konkretem Nutzerfeedback die Retrieval Performance deutlich verbessert werden konnte. Sie benutzten dazu einen Fragebogen, in dem der Suchende nach zusätzlichen Informationen über sein Suchziel gefragt wurde. Auf Basis dieser, über die ursprüngliche Query hinausgehenden Informationen wurde der Dokumentenspeicher erneut durchsucht und die Ergebnisse angepasst. Es wurde in dieser Studie offensichtlich, dass zwar zusätzliche Informationen notwendig sind (*query expansion*), dass aber mit deren Kenntnis die Ergebnisse des IR angepasst und verbessert werden können.

Zusammenfassend konzentriert sich die Forschung im Bereich des adaptiven Information Retrieval bisher vor allem auf die Verbesserung der Suchanfragen z.B. durch Berücksichtigung allgemeiner semantischer Relationen und/oder des Nutzerkontextes, im Sinne von vorangegangenen Suchanfragen und der Evaluation der Ergebnisse durch den Nutzer (in Form impliziten und/oder expliziten Feedback [Li, 2011]). Zum Einsatz kommen verschiedene Methoden und Techniken wie Entscheidungsbäume, Logik, Lernalgorithmen, Wissensrepräsentationsmodelle und andere.

Die Arbeiten zielen allesamt stark auf eine Verbesserung der Suchterme durch Berücksichtigung semantischer Relationen und/oder durch Anwendung von Logik und Lernverfahren ab. Kontext ist so vor allem auf den inhaltlichen Kontext einer Anfrage be-

zogen. Die genannten Untersuchungen belegen aber, dass die Bewertungen der Ergebnisse durch den Nutzer im adaptiven Verfahren besser ausfallen als bei den nicht adaptiven Suchvorgängen. Es bleibt allerdings offen, ob und wie auch über die inhaltlich hinausgehenden Faktoren der Kontext<sup>17</sup> des Nutzers konkret berücksichtigt werden kann.

Außerdem wird in dieser Arbeit nicht die situative Rezeption respektive Präsentation von *Dokumenten* sondern die von *Dokumentelementen* untersucht. Da die Untersuchungen dieser Arbeit vor allem auf die Verwendung verschiedener Inhaltstypen abzielen, wird also das als relevant ermittelte Dokument an sich in diese Elemente aufgebrochen. Das ist möglich, da zum einen in einem Wikipedia-Artikel die für das jeweilige Lemma semantisch relevanten Angaben enthalten sind (siehe dazu auch Kapitel 7.1.1) und zum zweiten immer der selbe Dokumenttyp zurückgeliefert wird (sowohl strukturell als auch intentionell), selbst wenn theoretisch mehr als ein relevanter Artikel zur Suchanfrage existiert. Ziel dieses Vorgehens ist es, über ein Re-Ranking von Dokumenten hinaus Aussagen zu einer situativen Informationsrezeption machen zu können.

### 7.2.3 Adaption mittels Berücksichtigung der Nutzerinteraktion

Ein weiterer (neben der Berücksichtigung semantischer Relationen) Weg, um Adaption zu ermöglichen, erfolgt durch Einbezug der Interaktion des Nutzers mit dem System [Kelly und Teevan, 2003; Joachims et al., 2007]. Durch Verwendung des in den Nutzeraktionen enthaltenen impliziten Feedbacks und dessen Berücksichtigung können Verbesserungen der Suchergebnisse aus Nutzersicht erzielt werden [White und Kelly, 2006].

Zum Thema des *Implicit Feedback* im Information Retrieval existieren bereits einige Untersuchungen und Erkenntnisse.

Jansen und Mc Neese [Jansen und McNeese, 2005] zeigen mit qualitativen Methoden, dass sich Muster in der Interaktion von Nutzern ermitteln lassen und eine automatische Suchassistenz dadurch verbessert werden kann. Morita und Shinoda [Morita und Shinoda, 1994] machten Zusammenhänge deutlich zwischen der expliziten Bewertung eines Textes durch den Nutzer und der Zeit, die er damit verbracht hat, diesen Text zu

---

<sup>17</sup> im Sinne des in dieser Arbeit definierten Kontextes.

lesen. Für das in dieser Arbeit intendierte Forschungsdesiderat ist das eine wichtige Erkenntnis, da so mittels der Interaktion auf die situative Relevanz eines Textes geschlossen werden kann (subjektive Relevanz eines Inhaltselements wird u.a. durch die Dauer der Berücksichtigung definiert, siehe dazu ab Kapitel 11.3).

Weitere Studien (z.B.[Claypool et al., 2001; Joachims et al., 2007]) konnten nachweisen, dass das Scroll- und Klickverhalten der Nutzer sowie ebenfalls deren mit den Inhalten verbrachte Zeit eindeutige Zusammenhänge zur expliziten Bewertung der verwendeten Inhalte aufweist. Damit konnten Rückschlüsse auf die nutzerspezifischen Interessen gezogen werden. Diese wiederum wurden erfolgreich dazu verwendet, um die Angebote zu verbessern. Auch die Blickbewegungen während der Suche [Buscher et al., 2008] liefern hilfreiche Hinweise in Form von implizitem Feedback, mit dessen Hilfe die Ergebnisse der Suche optimiert werden können.

In einer späteren Studie präzisieren Buscher et al. [Buscher et al., 2012] diese Ergebnisse. Die Autoren betonen, dass das *gaze-based Implicit Relevance Feedback* [Buscher et al., 2012] sehr gut geeignet ist, um die Websuche zu verbessern und die Suchergebnisse zu personalisieren.

Auch Untersuchungen von [Rayner, 1998; Nielsen, 2006] liefern Hinweise, dass durch die Analyse des Nutzerverhaltens Rückschlüsse auf deren Zufriedenheit mit den jeweiligen Inhalten getroffen werden können. Fox et al. [Fox et al., 2005] verwenden Mausklicks, die Verweildauer auf einer Seite sowie die Art und Weise, wie ein Nutzer die Suche beendete, als Indikatoren für seine Zufriedenheit.

Besonders interessant sind in diesem Zusammenhang auch die Arbeiten von Feild und Kollegen [Feild et al., 2010], die mit großer Genauigkeit durch das Klick-Verhalten der Probanden verschiedene Frustrationslevel bestimmen können. Die aktuelle Zufriedenheit des Nutzers während der Suche zu erkennen ist ein wichtiger Grundstein für eine adaptive Informationspräsentation.

Auch Navalpakkam und Churchill [Navalpakkam und Churchill, 2012] verwendeten Interaktionsdaten, um Aussagen über den Nutzer zu treffen: sie zeigen, dass die Posi-

tion und die Bewegungsmuster der Maus Hinweise liefert, wie gut der Nutzer mit einer Seite zurechtkommt und ob er gerade zufrieden ist.

Neben der Beobachtung der Interaktion mit dem Browser werden in dieser Arbeit auch die Blickbewegungen der Nutzer aufgezeichnet. Nach Stand der Forschung lassen sich in der Anwendung theoretisch auch die Blickbewegungen aus den Interaktionsdaten nachvollziehen, wie [Rodden et al., 2008] und [Guo und Agichtein, 2010-2] beschreiben. Guo und Agichtein identifizieren dazu drei verschiedene Typen einer Augen-Maus-Koordination: *Bookmarking* (Maus verweilt an bestimmten Stellen), *Incidental* (Mausbewegungen sind zufällig) und *Follow* (Mauszeiger folgt den Blickbewegungen des Nutzers). Vor allem im Fall *Bookmarking* kann die Mausposition für Implizites Feedback verwendet werden, da sie mit hoher Wahrscheinlichkeit einen für den Nutzer interessanten Inhalt markiert [Guo und Agichtein, 2010-2]. Bei diesem Typen ist es sogar möglich, aus den Maus-Events präferierte Inhaltselemente vorherzusagen. Auch der Typ *Follow* bietet einige Hinweise auf die Augenbewegungen des Nutzers und hinsichtlich relevanter Inhalte. In den in dieser Arbeit durchgeführten Studien konnte allerdings die Hand-Augen-Korrelation nicht bestätigt werden. Die Analyse der Browser-Interaktion zeigte keine eindeutigen Zusammenhänge mit den betrachteten Inhalten (siehe dazu auch Kapitel 11.4.4).

Huang et al. [Huang et al., 2012] unterscheiden etwas genauer zwischen den einzelnen Arten der *curser behaviour patterns* [Huang et al., 2012, S.1] wie lesen, scrollen, klicken und deren Zusammenhang mit den Blickbewegungen des Nutzers. Durch die Berücksichtigung der verschiedenen Muster gelang es den Autoren, die *gaze-position* anhand der Curser-Aktivitäten sehr genau zu approximieren. Für eine adaptive Informationspräsentation sind diese Erkenntnisse hilfreich, da so rückgeschlossen werden kann, welche Inhalte der Nutzer zur Problemlösung verwendet. Weitere Ergebnisse zur Hand-Augen-Korrelation sind zu finden bei [Guo und Agichtein, 2010-1].

Ein Zusammenhang zwischen *gaze position* und Curser-Bewegungen ist grundlegend, da im Falle eines adaptiven Interfaces auf die Browserinteraktion des Nutzers zurückgegriffen werden muss. Eyetracker stehen in der Regel dann nicht mehr zur Verfügung.

Weitere Informationen, die der Interaktion des Nutzers mit dem Browser entnommen werden können, zeigen Adafre und de Rijke [Adafre und Rijke, 2006]. Mit Hilfe von Logfiles zu Suchtasks in der Wikipedia gelang es ihnen, automatisch zwischen einer *directed* (also einer gezielten Suche nach etwas) und *undirected* (Suche mit unspezifischem Ziel) zu unterscheiden. Wenn die Hypothese, dass Nutzerpräferenzen abhängig von diesem Faktor unterschiedlich ausfallen bestätigt werden kann, so ist eine automatische Erkennung der gezielten bzw. ungerichteten Suche eine Grundlage, damit ein System entsprechend adaptiv reagieren kann.

Auch Moshfeghi und Jose [Moshfeghi und Jose, 2013] zeigen, dass Kenntnisse über die Ziele des Nutzers die Ergebnisse verbessern können. Dazu untersuchten sie die Charakteristika vier verschiedener Suchtasks, einmal die Suche nach Informationen an sich, weiter den Typ *information-refinding* – also die Situation, dass ein Nutzer bereits einmal gefunden Informationen wieder sucht, und zwei Typen von reinem *entertaining*-Charakter, etwa entsprechend den in dieser Arbeit verwendeten *casual-leisure* Situationen. Diese unterschieden sie noch zwischen *adjusting arousal* und *adjusting mood*. Die Tasks wurden den Nutzer in einer *simulated need situation* präsentiert [Borlund, 2000]<sup>18</sup>.

Die Testpersonen wurden sowohl in Fragebögen vor, während und nach der Studie befragt als auch deren Interaktion mit dem verwendeten System mitgeloggt. Unter Bezug auf das *mouse-gaze position alignment* entsprechend [Guo und Agichtein, 2010-1] konnte so ein Überblick gewonnen werden, wie lange die Nutzer jeweils die Ergebnisse betrachteten. Anhand von Fragebögen zeigen [Moshfeghi und Jose, 2013], welche Erfahrungen die Nutzer während der Suche gemacht haben, z.B. bezüglich ihrer Genauigkeit der Vorstellung der gesuchten Information, den Problemen, die Query zu formulieren oder der Vorstellung darüber, welche Informationen relevant sind. Es ist deutlich zu erkennen, dass sich diesbezüglich die Einschätzungen zwischen den verschiedenen Tasktypen teilweise stark unterscheiden. Die Erkenntnis, dass sich die Bewertungen der Ergebnisse zwischen diesen Klassen unterscheiden ist ein Indiz dafür, dass es sinnvoll ist, diese Aspekte eines Nutzungsszenarios bei der Informationspräsentation zu berücksichtigen.

---

<sup>18</sup> Siehe dazu auch 11.2.3.

In der Studie von Moshfeghi und Jose wurde ebenfalls deutlich, dass anhand der Interaktionsdaten statistisch signifikant unterschiedliche Verhaltensmuster zwischen den verschiedenen Aufgabentypen gefunden werden konnten. Deutlicher sind diese Ergebnisse noch, teilt man den Suchprozess in verschiedene Phasen [Moshfeghi und Jose, 2013]. Allerdings fehlen in dieser Studie Hinweise darauf, welche Inhaltstypen und -elemente in den verschiedenen Aufgaben bevorzugt wurden, bzw. wo die Unterschiede in der Bewertung der Inhalte durch die Nutzer bei den unterschiedlichen Aufgaben liegen.

Die besprochenen Arbeiten zeigen, dass zwischen dem Informationsbedürfnis, dem Suchverhalten und der Zufriedenheit des Nutzers mit den präsentierten Inhalten Rückschlüsse möglich sind und dass deren Berücksichtigung die Zufriedenheit der Nutzer im Information Retrieval verbessern kann.

Durch die Curser-Positionen und Aktivitäten können (teilweise) Rückschlüsse auf die betrachteten Inhalte gezogen werden. Weiterhin sind Zusammenhänge zwischen der mit den Inhalten verbrachten Zeit und der Zufriedenheit damit [Fox et al., 2005] sowie einer Identifikation von Unzufriedenheit [Feild et al., 2010] durch das Klickverhalten vorhanden.





## 8 Ergebnisse der Literatur und offene Fragestellungen

Zusammenfassend ist festzuhalten, dass gezeigt wurde, dass generell unterschiedliche Nutzungsszenarien vorhanden sind, die das Verhalten der Nutzer bei der Suche beeinflussen. Die Annahme einer subjektiven und situationsabhängigen Relevanzbewertung ist weiterhin begründet durch die Theorie des *cognitive viewpoint* und der Definition von Information als subjektives Konzept.

Die Ergebnisse der in der Forschungsliteratur vorhandenen Studien zum Einfluss bestimmter Kontextfaktoren auf das Nutzerverhalten wurden in einem Modell zusammengefasst und operationalisiert. Die positiven Effekte, die bereits mit adaptiven Systemen generell und im Information Retrieval speziell gezeigt werden konnten, sowie die Möglichkeiten, durch die Interaktion des Nutzers mit dem System konkrete Hinweise auf Zufriedenheit und Tätigkeit zu erhalten, sind eine plausible Basis dafür, dass beim Information Retrieval in der Wikipedia eine situations- und nutzerrelative *Informationspräsentation* die Nutzerzufriedenheit verbessern könnte.

Für den Anwendungsbereich der deutschsprachigen Wikipedia bleiben weiterhin Fragestellung offen, die im Folgenden empirisch untersucht werden, um diese Lücken in der bisher besprochenen Literatur zu ergänzen und die genannten Erkenntnisse und Theorien auf die Informationssuche in der Wikipedia anzuwenden.

Offene Fragestellungen:

1. Ist die deutschsprachige Wikipedia Gegenstand von sog. *casual-leisure* Suchszenarien?
2. Ist die Auswahl von Inhaltselementen subjektiv begründet und wenn ja, welches sind die Argumente für die Entscheidung für ein bestimmtes Objekt (aus einer Menge gleicher Medientypen)?
3. Unterscheiden sich die Nutzerpräferenzen und Informationsrezeption bei der Suche in der Deutschsprachigen Wikipedia abhängig vom Nutzungsszenario?

Den Fragestellungen wird nachgegangen, da die allgemeinen Erkenntnisse aus der Literatur nur dann auf die Wikipedia zutreffen, wenn diese erstens Gegenstand sowohl

---

von *work-based* [Borlund, 2000] als auch *casual-leisure* [Elsweiler et al., 2011] Szenarien ist, weiterhin gezeigt werden kann, dass Inhaltselemente nach subjektiven Kriterien und nicht (nur) thematischer Relevanz ausgewählt werden und drittens bei der Suche in der Wikipedia überhaupt abhängig vom Nutzungsszenario Unterschiede im Verhalten und den gewünschten Inhaltselementen bestehen.

## 9 Nutzungsszenarien in der Wikipedia

Mit den folgenden Untersuchungen und Darstellungen werden die Gründe aufgezeigt, die die Annahme bestärken, dass die Wikipedia neben ihrer Funktion als Nachschlagewerk (sog. *work-based*-Szenarien, siehe [Lim, 2009]) auch zur Freizeitsuche herangezogen wird.

### 9.1 *Work-based* Nutzung der Wikipedia

Head und Eisenberg [Head und Eisenberg, 2010] zeigen in ihrer Untersuchung, dass College-Studenten an sieben verschiedenen Einrichtungen (insgesamt 84 Teilnehmer) die Wikipedia regelmäßig für sog. *work-based* Szenarien, also durch einen bestimmten Informationsbedarf motivierte Suchsituationen, verwenden. Nur 9% der Befragten verwendeten nie die Wikipedia, der Rest mindestens gelegentlich und 30% sogar *always* [Head und Eisenberg, 2010]. Dieser Studie konnte auch entnommen werden, dass für die häufige Nutzung vor allem auch die Breite der Inhalte und die Aktualität der Wikipedia ausschlaggebend sind.

Auch Lim [Lim, 2009] berichtet Ähnliches: Etwa ein Drittel von 134 befragten Studenten verwendeten die Wikipedia *for academic purposes* [Lim, 2009]. Dabei standen vor allem das Fakten Retrieval (entsprechend *look up* nach [Marchionini, 2006]) und ausführlichere Hintergrundinformationen (entsprechend *learn* nach [Marchionini, 2006]) im Fokus.

Ob eine Nutzung der Wikipedia in den oben beschriebenen *casual-leisure* Szenarien stattfindet, ist ungeklärt, da die Studien hier widersprüchlich sind.

### 9.2 Freizeitsuche in der Wikipedia

Die genannten Studien [Head und Eisenberg, 2010; Lim, 2009] zeigen eine Nutzung der Wikipedia für *work-based* Absichten. Allerdings sind diese Studien widersprüchlich, was die Nutzung der Wikipedia für *non-work-based* Szenarien betrifft.

In einer eigenen Studie (siehe ab Kapitel 11) gaben 18 von 29 Nutzern (für genauere demographische Angaben siehe 11.3 ) an, dass sie die WP oft oder sehr oft auch zum

Zeitvertreib benutzten (vier bzw. fünf von fünf möglichen Punkten, die zu vergeben waren) und lediglich acht der Testpersonen bewerteten die Bedeutung der Suche in der Wikipedia als Freizeitbeschäftigung mit zwei oder einem Punkt (also als unwichtig). Diese Ergebnisse bieten einen guten Anhaltspunkt, allerdings handelt es sich bei der Studie um eine relativ kleine und nicht repräsentative Stichprobe.

Zusätzlich zu diesen Anhaltspunkten wird deshalb untersucht, ob Aufbau und Beschaffenheit sowie die Themenverteilung der Inhalte in der deutschsprachigen Wikipedia Hinweise auf eine Beschäftigung zu Freizeitzwecken enthalten.

### 9.3 Geschichte und Entwicklung der Wikipedia

Die Wikipedia wurde am 15. Januar 2001 gegründet. Dahinter steckte kein kommerzielles Projekt, sondern ursprünglich die Idee, das Wissen der Welt zu „befreien“ und in vielen verschiedenen Sprachen zur Verfügung zu stellen [Stegbauer, 2012]. Dazu bemüht sich die Community, eine qualitativ hochwertige Enzyklopädie zu verfassen. Die größte Version ist dabei im Moment (Stand Juli 2013) die englischsprachige Version mit über vier Millionen Artikeln, gefolgt von der deutschsprachigen Wikipedia mit über 1,5 Millionen Artikeln.

Die Inhalte der Wikipedia werden von einer weitgehend anonymen Autorengruppe erstellt und nicht wie bei traditionellen Enzyklopädien üblich durch eine Redaktion organisiert. Grundsätzlich ist die Mitarbeit jedem zu jeder Zeit möglich, in der Realität allerdings gestaltet sich das oft eher schwierig. Eigentlich entscheidet zwar die „Gemeinschaft“ – sprich, in Diskursen werden strittige Frage z.B. um die Relevanz eines Artikelgegenstands etc. geklärt – de facto aber wird die Wikipedia dominiert von einer Gruppe von Aktivisten, die Richtung und Dynamik des Projekts bestimmen [Stegbauer, 2012].

Nichtsdestotrotz bietet die Wikipedia eine umfangreiche Sammlung an Artikeln zu einer hohen Themenvielfalt. Experten bestimmter Fachgebiete betreuen z.B. sog. *Themenportale* und stellen ihr Fachwissen zur Verfügung. So entstehen auch sehr fundierte, umfangreiche und mit Expertenwissen versehene Artikel mit großer Detailtiefe.

Tatsächlich schafft das Kollektiv der Wikipedia-Autoren eine gigantische Wissenssammlung von teilweise (auch vom Bearbeitungsstand eines Artikels abhängig) hoher Qualität (siehe auch 9.6).

Für die Autoren und Mitarbeiter ist die Arbeit an der Wikipedia ein schönes Hobby, sie sind teilweise motiviert durch die Ideologie des „befreiten Wissens“ oder aber dem Spaß daran, ein wirklich gutes Produkt zu schaffen [Stegbauer, 2012].

## 9.4 Inhaltstiefe, Details und Sprache in der Wikipedia

Sprache und der Grad an Detailliertheit variieren zwischen den Artikeln. Das liegt vor allem am „Alter“ und dem Bearbeitungszustand eines Artikels. Abgeschlossene Artikel werden zur Bearbeitung gesperrt und haben in der Regel ein vergleichbares Niveau. Grundsätzlich haben sich der Sprachstil und die Terminologie anderer, traditioneller Universallexika durchgesetzt, dennoch sind diese auch abhängig vom Autor/den Autoren eines Artikels.

Es lässt sich feststellen, dass sprachliche Fehler bzw. auch grammatikalische oder Interpunktions-Fehler von der Community relativ schnell korrigiert werden, was insgesamt ein hohes sprachliches Niveau schafft (siehe auch Kapitel Informationsqualität).

Stärker variiert die Inhaltstiefe der Artikel. Gerade etwa bei philosophischen Themen arbeiten oft sehr versierte Autoren enthusiastisch an Artikeln über Theorien oder Lehren mit. Dabei entstehen leicht Diskussionen darüber, bis zu welchem Komplexitätsgrad Sachverhalte in die Artikel aufgenommen werden sollen. Für Fachexperten ist hierbei natürlich die Schwelle deutlich höher als für den „durchschnittlichen“, eher universell versierten Wikipedia-Autor. Theoretisch wird ein Artikel in erster Linie höherwertiger, je bessere, detailliertere und umfangreichere Informationen er enthält. Aber das gilt gerade in der Wikipedia nicht absolut. Denn erstens sollen die Artikel nur einen Überblick über das jeweilige Lemma geben, da es sich um Enzyklopädie-Artikel handelt, von deren Inhalt der Leser eine Verständlichkeit ohne Vorwissen und die Vermittlung von Grundlegendem erwarte. Zweitens hat jede Disziplin eine Fachsprache und verwendet Terminologien unterschiedlich, was durchaus zu Verständnisproblemen bei Fachfremden führen kann. Ein zu detailliertes Niveau der Inhalte birgt daher das Risiko, dass die Relevanz und der Nutzen für den Leser aufgrund seiner Fach-

fremde sinkt. Das ist ein weiterer Grund, weshalb eine *Auswahl* an Inhalten entsprechend dem Nutzer (und seiner kognitiven Struktur) sinnvoll scheint. Es wird bei der Wikipedia viel Fachwissen eingebracht, das nicht für alle Leser verständlich, allerdings für themenerfahrene Nutzer erfreulich ist.

Grundsätzlich lässt sich festhalten, dass die Inhaltstiefe und die Details zwischen den Artikeln variieren, abhängig davon, ob es ausreichend Experten für Details gibt und ob Details erwünscht sind. Für die Untersuchungen in dieser Arbeit sind die Artikel detailliert und umfangreich genug, sie sind gut und vergleichbar strukturiert und abwechslungsreich gestaltet. Sprachliches Niveau, Detailtiefe und Umfang der Artikel sind entsprechend hoch, so dass auch eine Beschäftigung mit den Artikeln in der Freizeit grundsätzlich vorstellbar ist.

## 9.5 Inhaltsstrukturen

Als Inhalt der deutschsprachigen Wikipedia sind zentral die Artikelgegenstände zu bezeichnen. Die Artikel selbst setzen sich wiederum aus einzelnen Elementen zusammen wie etwa Text, Bild, Überschriften etc. Außerdem gehören zum Inhalt der Wikipedia weiterhin Diskussionsseiten (auf denen die Autoren sich über die Inhalte der Artikel und eventuell strittige Fragen austauschen können), Hilfeseiten mit Richtlinien und Tipps für Autoren, Listen, Versionsgeschichten für jeden Artikel, Autorensseiten, auf denen die Mitglieder persönliche Informationen zur Verfügung stellen und einiges mehr.

In dieser Arbeit wird konkret auf diese Inhaltselemente und ihre situationsabhängige Eignung Bezug genommen. Es wird nur mit den vorhandenen Strukturen gearbeitet und es werden keine Vorschläge für neue Inhalte gemacht, da die vorhandenen strukturellen Maßnahmen der Wikipedia bereits den Aspekt des unterschiedlich umfangreichen Informationsbedürfnisses teilweise berücksichtigen<sup>19</sup>. Die nachfolgende Beschreibung der Inhaltselemente entspricht auch den Gegenständen der später verwendeten

---

<sup>19</sup> Teilweise deshalb, da immer alle Elemente präsentiert werden, sowohl Übersichten und Zusammenfassungen als auch ausführliche Texte und Illustrationen.

Videolabels (siehe ab Kapitel 11) bzw. dem Wertebereich der betrachteten Elemente im Kontextmodell (siehe ab Kapitel 6).

Ein vollständiger Artikel in der deutschsprachigen Wikipedia beginnt mit der Überschrift, die den Artikelgegenstand enthält. Darunter folgt eine allgemeine, zusammenfassende Einleitung zum Thema. Rechts oben kann sich eine sog. Infobox befinden, die meist ein Foto oder eine Infografik enthält sowie tabellenartig zusammengefasste Fakten über den Artikelgegenstand (z.B. bei einer Stadt oder einem Land Einwohnerzahlen und Fläche, dazu das Stadtwappen, siehe Abbildung 9-1).

Wappen	Deutschlandkarte
	
Basisdaten	
Bundesland:	<a href="#">Bayern</a>
Regierungsbezirk:	<a href="#">Oberpfalz</a>
Höhe:	343 m ü. NN
Fläche:	80,76 km²
Einwohner:	136.577 <i>(31. Dez. 2011)</i> <sup>[1]</sup>
Bevölkerungsdichte:	1691 Einwohner je km²
Postleitzahlen:	93047–93059
Vorwahl:	0941
Kfz-Kennzeichen:	R
Gemeindeschlüssel:	09 3 62 000
Stadtgliederung:	18 <a href="#">Stadtbezirke</a>
Adresse der Stadtverwaltung:	Rathausplatz 1 93047 Regensburg
Webpräsenz:	<a href="http://www.regensburg.de">www.regensburg.de</a> 
Oberbürgermeister:	<a href="#">Hans Schaidinger</a> (CSU)

Abbildung 9-1 Infobox des Wikipedia-Artikel über Regensburg, abgerufen Mai 2013 von <http://de.wikipedia.org/wiki/Regensburg>

Der Artikel-Text selbst wird gegliedert durch Über- und Unterüberschriften, die **fett** und in größerer Schrift hervorgehoben werden. Bilder, Grafiken, Tabellen sowie Video und Audiomaterial können an jeder Stelle im Artikel eingefügt werden. Bilder befinden sich in der Regel rechts oder links außen, es gibt aber auch die Möglichkeiten, Bildergalerien zu erstellen wenn mehrere Bilder an einer Stelle eingefügt werden sollen.

Zwischen dem Artikel-Text und der Einleitung befindet sich ein Inhaltsverzeichnis, das durch Aggregieren der Überschriften entsteht und mit dessen Hilfe zu den einzelnen Abschnitten des Artikels navigiert werden kann.

Am Ende eines Artikels befinden sich die verwendeten Nachweise so wie weiterführende Links (z.B. weitere Literatur, Bilder, Webseiten). Darunter stehen die Informationen zum Artikelgegenstand, z.B. zu welcher Inhaltskategorie im Wikipedia-internen Kategoriensystem der Artikel verortet ist, wann er zuletzt geändert wurde, ob der Artikel in die Liste der guten oder exzellenten Artikel aufgenommen wurde, welchen Lizenzen er eventuell unterliegt und ein Hinweis auf die Nutzungsbedingungen der Wikipedia.

Die Länge der Artikel in der Wikipedia variiert, ebenso die Ausstattung mit nicht textuellen Elementen.

## 9.6 Informationsqualität in der Wikipedia

Beim selektiven Vorgehen der Nutzer spielt auch die Qualität der angebotenen Inhaltselemente eine Rolle [Wang und Soergel, 1998]. Es ist damit vorstellbar, dass selbst, wenn ein Fließtext präferiert würde, z.B. eine Tabelle lieber betrachtet wird, sollte die Qualität des Fließtextes mangelhaft sein. Die Qualität der Inhalte in der Wikipedia wird deshalb hier besprochen.

Auch in der Wikipedia befinden sich nicht korrekte Angaben, sogar bewusste Falschmeldungen. Durch die transparente Bearbeitung allerdings lässt sich das teilweise kontrollieren, denn Veränderungen sind immer kenntlich gemacht und müssen auch in der Regel durch Administratoren freigegeben werden. Das schützt allerdings nicht vor unwissentlich platzierten fehlerhaften Informationen.

Grundsätzlich sind Bedenken zur Qualität nicht von der Hand zu weisen, denn es ist fraglich, wie eine Qualitätskontrolle stattfinden kann, wenn jeder alles verändern kann (mit Einschränkungen, Artikel können gesperrt werden wenn z.B. ein sog. *edit-war*<sup>20</sup>

---

<sup>20</sup>*Edit-war* bedeutet, dass Autoren gegenseitig die Änderungen des jeweils anderen sofort wieder rückgängig machen oder überschreiben, meistens motiviert durch Fehlen eines Konsens über bestimmte Streitpunkte.



droht oder aber sie werden nicht weiter für die Allgemeinheit zur Bearbeitung freigegeben da sie bereits eine sehr hohe Qualitätsstufe erreicht haben).

Es gab bereits auch Fälle, bei denen Politiker oder große Konzerne bewusst Artikel zu ihrem Vorteil oder jemand anderes Nachteil verändert haben [Grzanna et al., 2011]. Durch die Funktionsweise der Wikipedia (siehe Kapitel 9.3) werden solche Vorkommnisse allerdings von der Gemeinschaft in der Regel zeitnah aufgedeckt. Generell stehen in der Wikipedia die Information und ihre Qualität immer im Spannungsfeld zwischen den verschiedenen Autoren. Das ist zunächst bewusst gewollt, um dadurch – theoretisch – eine Optimierung zu erreichen. Die Idee dahinter ist, dass sich aus den vielen Meinungen und Beiträgen der mittlerweile sehr großen Menge der Wikipedianer die jeweils besten Beiträge durchsetzen.

In der Theorie bedeutet das, dass festgelegt ist, was „richtig“ oder „das Beste“ ist und das System der kollaborativen Arbeitsweise sorgt dafür, dass die entsprechenden Informationen bestehen bleiben.

Tatsächlich aber ist diese Frage komplexer, da es für vieles bewusst *keine* konkreten Regeln gibt, da die Wikipedia so frei wie möglich, sprich mit einem Minimum an Regeln auskommen möchte. Für manche Punkte werden Richtlinien auf der Seite aufgeführt, die wiederum im Konsens und der Eigenregie der Usergemeinschaft entstanden sind. Bei anderen Punkten allerdings fehlt dagegen ein Konsens, der als Regel, Vorgabe oder Maßstab für die Optimierung der Beiträge herangezogen werden könnte.

Die Auffassung von hoher Qualität ist bei den Autoren in der Wikipedia bisweilen unterschiedlich. Dabei erstrecken sich Streitfragen nicht nur auf die faktische Richtigkeit von Informationen, sondern es geht auch um ästhetische, ethische, wissenschaftliche, weltanschauliche und grundsätzliche Dinge. Auch wenn die Regel des „Neutralen Standpunktes“<sup>21</sup> gilt, so gibt es hier bisweilen keinen Konsens. Das liegt in erster Linie an der Heterogenität der Autoren und den damit zahllosen Standpunkten, Geschmacksfragen, Bildungsständen und ähnlichem. So sind auch die Ziele und Interessen der zahlreichen Autoren und Leser der WP oft nicht die gleichen. Wie [Wang und Soergel, 1998] beschreiben, spielen auch bei der Auswahl bzw. bei der Entscheidung

---

<sup>21</sup> Ein proprietäres Konzept der Wikipedia zur Vermeidung von Grundsatzdiskussionen und Wertungen in den Artikeln, siehe dazu: Wikipedia: Neutraler Standpunkt: [[http://de.wikipedia.org/wiki/Wikipedia:Neutraler\\_Standpunkt](http://de.wikipedia.org/wiki/Wikipedia:Neutraler_Standpunkt)] (Abgerufen 5.Juni 2014).

des Nutzers für oder gegen einen Inhalt diese Kriterien eine Rolle. Somit kann ein Inhalt auch aus z.B. ethischen Gründen abgelehnt werden.

Generell ist die Qualität der Wikipedia Artikel vergleichbar mit denen einer redaktionell erstellten Enzyklopädie, wie eine Studie der *Nature* [Giles, 2005] zeigt. Hier wurden 50 Artikel einer Zufallsstichprobe zweier Enzyklopädien aus einem breiten naturwissenschaftlichen Themenfeld (englischsprachige Wikipedia und *Enzyklopedia Britannica*) ausgewertet [Giles, 2005]. Die Artikel wurden von Fachexperten bewertet, die jeweils nicht wussten, aus welchem Werk der jeweilige Artikel stammte. Es zeigte sich, dass die Qualität zwar in einigen Artikel stärker schwankte, generell aber nicht sehr unterschiedlich war.

Einen guten Ein- und Überblick über die Qualitätsforschung an der Wikipedia gibt Hammwöhner [Hammwöhner, 2007]. Er zeigt anhand verschiedener Beispiele [Giles, 2005; Stvilia et al., 2005a; Wiegand, 2007, Stvilia et al., 2005b], wo methodische Schwierigkeiten bei der Qualitätsmessung an der Wikipedia auftreten. Diese Schwierigkeiten reichen von der Auswahl der Qualitätskriterien bis hin zur Auswahl des zu bewertenden Umfangs. So beziehen sich die einen Studien nur auf die inhaltliche Qualität einiger Artikel, während andere auch die Softwareumgebung und die Struktur der Enzyklopädie berücksichtigen. Da die Qualität nach [Wang und Soergel, 1998] ein Kriterium ist, das bei der Evaluation der Inhalte durch den Nutzer eine Rolle spielt, ist die Qualitätssicherung ein wichtiger Aspekt.

[Hammwöhner, 2007] testete die Qualität der Inhalte der deutschen Wikipedia im Vergleich mit der einer Printausgabe der Brockhaus Enzyklopädie. Dabei wurden anhand einer Stichprobe die verschiedenen Artikel auf Umfang, Vollständigkeit, Quantität und Qualität der Quellenangaben sowie formale Korrektheit untersucht. Das Ergebnis fiel je nach Kategorie zwar unterschiedlich aus, in Summe aber waren die Artikel in beiden Enzyklopädien von ähnlicher Qualität. Im Vergleich zu einer früheren Studie von Schlieker im Jahr 2005 [Schlieker, 2005] war aber vor allem ein Fortschritt seitens der Wikipedia bemerkenswert.

Zusammenfassend ist somit festzuhalten, dass die Inhalte der Wikipedia in Fragen der Qualität, der Inhaltstiefe, der Struktur und dem (möglichen) Umfang der Artikel theoretisch geeignet sind, um auch Gegenstand explorativen Stöberns [Adafre und Rijke, 2006] und/oder einer Freizeitsuche zu sein.

## 9.7 Verteilung der Inhalte auf die Wissensdomänen

Ein weiteres Detail, das zur Vermutung einer Freizeitnutzung der Wikipedia geführt hat, ist neben den bisher beschriebenen Beobachtungen und Untersuchungen die Annahme, dass sich die Verteilung der Themen in der Wikipedia im Vergleich zu einer traditionellen Enzyklopädie verändert hat. Es besteht der Verdacht, dass der Anteil der Artikel aus Themen des „Freizeitbereichs“ (Sport, Unterhaltung, Alltag und Gesellschaft) im Vergleich zu einer traditionellen Enzyklopädie zugenommen hat.

Eine Untersuchung, die Spoerri [Spoerri, 2007] zwischen September 2006 und Januar 2007 durchführte, zeigte, dass 43% der monatlichen 100 meistbesuchten Seiten der Wikipedia in den Bereich „Entertainment“ [Spoerri, 2007] fallen (siehe Abbildung 9-2).

<b>Categories</b>	<b>Total (230)</b>	<b>Percentage of total</b>	<b>5 (39)</b>	<b>4 (15)</b>	<b>3 (19)</b>	<b>2 (31)</b>	<b>1 (126)</b>
Entertainment	100	43%	8	7	11	17	57
Politics + History	34	15%	5		2	6	21
Geography	28	12%	8	3	2	2	13
Sexuality	23	10%	11	2		3	7
Science	13	6%			2		11
Computers	12	5%	6	2			4
Arts	5	2%		1		1	3
Religion	5	2%				2	3
Holidays	5	2%			1		4
Current events	3	1%	1		1		1
Drugs	2	1%					2

Abbildung 9-2 Monatliche Top 100 Wikipedia-Artikel Sept. 2006 - Januar 2007 aus [Spoerri, 2007]

Eine andere Studie dagegen [Lim, 2009], die bei der Zielgruppe Studenten durchgeführt wurde, bestätigte dieses Ergebnis nicht: hier gaben 61,9% der Nutzer an, die Wikipedia nicht für die Recherche zu Themen aus dem Entertainment-Bereich zu benutzen.

zen [Lim, 2009]. Da davon auszugehen ist, dass die verbleibenden knapp 38,1 % der Nutzer, die die Wikipedia zu Entertainment-Themen durchsuchen, dies nicht *ausschließlich* tun, sinkt der Prozentsatz der Studie von [Lim, 2009] zu Folge deutlich im Vergleich zu den Ergebnissen von [Spoerri, 2007].

Die beiden Studien zeigen, dass einerseits „Freizeitthemen“ in der Wikipedia populär sind, aber andererseits nicht generell von den Nutzern als wichtige Themen bzw. als vorstellbares Nutzungsszenario bewertet werden. Deshalb wird im Folgenden die generelle Verteilung der Inhalte in der deutschsprachigen Wikipedia auf die verschiedenen Themengebiete untersucht. Da die Themen auch von der Community stark mitbestimmt werden, ist die Themenverteilung als Hinweis für die entsprechende Nutzung bzw. die Ansprüche der Community (siehe dazu auch 9.3) zu sehen.

Für diese Untersuchung muss ein Schema zur eindeutigen Sortierung der Inhalte vorliegen. Diese Zuordnungsmöglichkeit wird auch für das erstellte Modell (siehe ab Kapitel 6) benötigt, um die Nutzungsszenarien konkret beschreiben zu können. Das Sortierungssystem wird auf Basis des vorhandenen Kategoriensystems der deutschsprachigen Wikipedia entwickelt. Es wird dargelegt, weshalb dieses System nicht direkt übernommen werden konnte, die zusätzlichen, aufgabenspezifischen Anforderungen an das zu entwickelnde System werden erläutert.

Ziel ist es, durch einen Stichproben-Vergleich aus der Brockhaus Enzyklopädie<sup>22</sup> und der deutschen Wikipedia (Stand Juni-August 2013), die Verteilung der Inhalte auf die verschiedenen Bereiche zu untersuchen und diese Verteilung zwischen den beiden Werken<sup>23</sup> zu vergleichen. Dabei liegt der Fokus vor allem darauf, welchen Anteil die Themen aus dem Bereich Freizeit/Entertainment einnehmen.

Eine Veränderung der Verteilung der Inhalte hin zu „Freizeitthemen“ wird als weiterer Hinweis darauf gesehen, dass die Wikipedia auch Gegenstand von sog. *casual-leisure* Szenarien nach [Elsweiler et al., 2011] ist.

---

<sup>22</sup> 17. Auflage, 1966-1976

<sup>23</sup> Es musste die Printversion der BE verwendet werden, da es nicht möglich war, Zufallsstichproben aus der Online-Version zu bekommen.

### 9.7.1 Bestehende Ordnung in der Wikipedia

In der Wikipedia existiert bereits ein umfassendes System zur sachsystematischen Verortung der Artikel. Es wurde von der Community 2004 in Betrieb genommen. Dieses System umfasst mehrere 1000 Kategorien, ist oft zehn und mehr Kategorien tief und lässt auch Mehrfachzuordnungen zu (z.B.: Lemma „Katze“: Katzen/Hauskatze/Haustier<sup>24</sup>). Da das bestehende Kategoriensystem sehr groß ist, Mehrfachzuordnungen zulässt, keine eindeutigen Zuweisungsregeln hat und zudem beliebig erweiterbar ist, ist es für die hier durchzuführende Untersuchung nicht geeignet. Allerdings soll es bei der Erstellung des zur Analyse und zur Verwendung im Kontextmodell erstellten Systems berücksichtigt werden, da der Geltungsbereich gleich bleibt und so das Kategoriensystem der Wikipedia für die hier gewünschte Klassifikation der Inhalte eine gute Grundlage liefert.

Die Erstellung eines Ordnungssystems ist eigentlich ein komplexer und aufwändiger Prozess. Für das Kategoriensystem in der Wikipedia trifft diese Vorgehensweise nicht zu. Wie auch für die Erstellung der Inhalte, gilt hier das Prinzip des kollektiven Wissens, zu dem jeder etwas beitragen kann. Jeder kann Kategorien hinzufügen und diese jeweils anderen Kategorien unterordnen. Jede Kategorie kann beliebig vielen anderen Kategorien zugordnet werden.

Die Artikel werden den Kategorien ihrem Inhalt entsprechend zugewiesen. Es gilt, dass Oberkategorien allgemeinere Begriffe repräsentieren als Unterkategorien. Ansonsten gibt es keine genauen Zuweisungsregeln außer den Grundsätzen, dass ein Lemma entweder zum Fachgebiet gehören muss, Teil (von) oder zeitlich oder örtlich zugehörig ist<sup>25</sup>. Jeder Artikel kann mehrfach zugeordnet werden, dabei sollte allerdings eine Seite nicht gleichzeitig einer Kategorie *und* einer ihrer Unterkategorien zugeordnet werden<sup>26</sup>. Kategorien sollten, wenn sie mehr als 200 Einträge haben, in Unterkategorien aufgeteilt werden. Die jeweilige Aufteilung der Kategorien wird den Fachbereichen überlassen – nur die allgemeinen Oberkategorien sind für alle Themen- und Fachbereiche gleich. In

---

<sup>24</sup> Siehe Wikipedia: Katze [<http://de.wikipedia.org/wiki/Katze>], 19.Juli 2010.

<sup>25</sup> Zum Kategoriensystem in der Wikipedia und die Zuordnungsregeln siehe [<http://de.wikipedia.org/wiki/Wikipedia:Kategorien>] (zuletzt abgerufen 26.9.2013) und [<http://de.wikipedia.org/wiki/Hilfe:Kategorien>] (zuletzt abgerufen 26.9.2013).

<sup>26</sup> Ebda.

einer Beschreibung, die jede Kategorie auf ihrer Kategorienseite erhält, werden der Zweck der Kategorie und die Abgrenzungskriterien zu den anderen Kategorien festgehalten. Anders als bei Kategoriensystemen für spezielle, abgegrenzte Fachgebiete, muss das System in der Wikipedia auch sehr allgemeine Bereiche umfassen.

Die Beziehungen zwischen den Kategorien selbst sind nicht weiter definiert. Es werden zwar Zuordnungsregeln für die Unterkategorien gegeben, allerdings nicht für alle. Außerdem gibt es auch eindeutige Anweisungen für Mehrfachzuordnungen, zum Beispiel sollen Artikel, die zu „Räumliche Systematik: Inselgruppe als Thema: Kanarische Inseln“ gehören auch zum jeweiligen Staat zugeordnet werden<sup>27</sup>. Auf den jeweiligen Seiten der Unterkategorien befinden sich dann wiederum Zuweisungsregeln. Die weitere Untergliederung ist relativ kurz, klickt man sich, je nach Inselgruppe, durch die einzelnen Kategorien, so kommt man nach weiteren drei (Insel Panarea, gehört zu den Liparischen Inseln) bis fünf (Britische Jungferninseln) zum Artikel. Über die Konsistenz der Zuordnung lässt sich durch die große Zahl der Kategorien ad hoc wenig sagen.

Für das Anliegen dieser Arbeit ist eine so feine Untergliederung aus oben genannten Gründen weder sinnvoll noch machbar. Daher gibt es vor allem für die hierarchisch niedrigeren Unterkategorien des Kategoriensystems der Wikipedia für die hier intendierten Untersuchungen keine weitere Verwendung.

### ***Statistiken zum Kategoriensystem in der Wikipedia***

Um die Verwendungsmöglichkeiten des Kategoriensystems der Wikipedia für die hier geplanten Untersuchungen genauer zu betrachten wurden weitere Aspekte untersucht. Die Kategorien des Systems sind alphabetisch geordnet den Wikipedia-Seiten zu entnehmen. Am 3. August 2010 hatte z.B. der Buchstabe B ca. 9500 Kategorien. Das zeigt deutlich, dass das System auf Grund seiner Größe schwierig zu handhaben ist. Abgesehen davon wären keine einheitlichen Schlussfolgerungen für ein System von der durchsuchten Domain auf die Präferenzen der Nutzer möglich, es sei denn, man abstrahierte wieder von den detaillierten Domains bzw. fasste diese in Klassen zusammen, was so gesehen ja das Ziel des hier erstellten Kategoriensystems ist.

---

<sup>27</sup> Ebda.

Die Kategorien mit den meisten Zuordnungen sind wiederum recht allgemeiner Natur (bereinigt von Kategorien, die nicht inhaltlicher sondern organisatorischer Natur sind):

*Tabelle 9-1 Kategorien mit den meisten Zuordnungen in der Deutschsprachigen Wikipedia*

1	Mann [287.329 Einträge]
2	Deutscher [88.665 Einträge]
3	Frau [48.269 Einträge]
4	Autor [31.351 Einträge]
5	US-Amerikaner [29.156 Einträge]
6	Schauspieler [21.628 Einträge]
7	Österreicher [12.422 Einträge]
8	Literatur [Deutsch][9.281 Einträge]
9	Franzose [8.609 Einträge]
10	US-amerikanischer Film [7.923 Einträge]
11	Brite [7.800 Einträge]
12	US-amerikanischer Musiker [6.367 Einträge]
13	Schweizer [6.148 Einträge]
14	SPD-Mitglied [6.105 Einträge]
15	Literarisches Werk [6.056 Einträge]
16	Roman, Epik [5.982 Einträge]
17	Geboren im 20. Jahrhundert [5.568 Einträge]
18	Fluss in Europa [5.276 Einträge]
19	Pkw-Modell [5.263 Einträge]
20	Italiener [5.223 Einträge]
21	Fußballspieler [Deutschland][4.577 Einträge]
22	Literatur [Englisch][4.427 Einträge]

Da es sich hierbei um die am öftesten vergebenen Kategorien handelt, lässt dies bereits erste Rückschlüsse auf die inhaltlichen Schwerpunkte der Wikipedia zu (siehe z.B. Rang 21: Fußballspieler (Deutschland)). Allerdings ist das nur bedingt aussagekräftig. Betrachtet man die absoluten Zahlen der Einträge in den meistbenutzten Kategorien, so ist bereits ab Rang 8 die Population unter 10.000 Einträgen, was in Relation zu der gesamten Artikelanzahl (Stand: 4. August 2010, 9:46 Uhr: 1.090.454 Artikel) relativ wenig ist.

Die ersten drei meistbenutzten Kategorien („Mann“, „Deutscher“, „Frau“) haben zwar deutlich mehr Einträge, sind aber wiederum sehr allgemein und vermutlich wurden viele Artikel sowohl der Kategorie „Mann“ als auch „Deutscher“ zugeordnet. Was allerdings eine gewisse Aussagekraft hat ist, dass Kategorie „Mann“ + die Kategorie „Frau“ (hier gibt es vermutlich keine doppelten Zuordnungen) gesamt 335598<sup>28</sup> Einträge hat, was eine große Abdeckung der Kategorie „Personen“ ergibt, wie die späteren Stichproben auch bestätigen.

---

<sup>28</sup> Stand: 4. August 2010

Da es zum Zeitpunkt der Untersuchung nur 26 nicht kategorisierte Artikel gab<sup>29</sup>, kann man diese Artikelzahl realistisch ins Verhältnis setzen zur Gesamtartikelanzahl, was immerhin einen Prozentsatz von 30,8 % für die Artikel über Personen ergibt. Um Aussagen über weitere Klassen von Artikelgegenständen machen zu können müssen diese ebenso eindeutigen Oberklassen wie „Person“ zugeordnet werden können. Darüber hinaus ist weiterhin von Interesse, auch den thematischen Bereich, dem die Artikelgegenstände entstammen, zu erfassen, also z.B. Sport, Wissenschaft oder Kunst/Kultur.

### 9.7.2 Anforderungen an das Kategoriensystem

Für den beschriebenen Zweck wird ein überschaubares Ordnungssystem benötigt. Es soll höchstens so groß sein, dass ein Benutzer des Systems jederzeit alle Kategorien überblicken kann. Wenn ein Benutzer vor der Aufgabe steht, ein Lemma einzuordnen und mit dem System noch nicht vertraut ist, dann muss es ihm möglich sein, für jedes Lemma das Ordnungssystem in einer angemessenen Zeit einmal zu durchlaufen, um die richtige Kategorie zu finden. Das System soll deshalb maximal vier Ebenen haben, bzw. ein Lemma soll nicht mehr als vier Attribute zugordnet bekommen. Das von [Spoerri, 2007] verwendete System konnte hier nicht herangezogen werden, da in der vorliegenden Arbeit nicht nur der Bereich, sondern zusätzlich auch der Artikelgegenstand (also z.B. Person oder Ding) erfasst werden sollen. Das System soll auch bei einer gezielten Untersuchung der Themenabhängigen Informationsrezeption bei der Wikipedia eingesetzt werden können [siehe dazu Vogel, 2013]

Für das Ordnungssystem werden weiterhin eindeutige Zuweisungsregeln benötigt. Eine Mehrfachzuordnung darf nicht möglich sein.

Ohne Bezugsrahmen und Absichtserklärung können keine Aussagen getroffen werden, die objektiv und sinnvoll interpretiert und verwendet werden können. Deshalb sind vor der Erstellung des Ordnungssystems folgendes festgelegt:

**Sachgebiet:** Im Falle der Wikipedia ist das Sachgebiet das „Wissen der Welt“, das die Wikipedia zu sammeln gedenkt.

---

<sup>29</sup> Quelle: [[http://de.wikipedia.org/wiki/Spezial:Nicht\\_kategorisierte\\_Seiten](http://de.wikipedia.org/wiki/Spezial:Nicht_kategorisierte_Seiten)], 4. August 2010.



**Dokumentationseinheit:** Die hier zu ordnenden Einheiten sind die Enzyklopädie - Artikel bzw. deren Beschreibungsgegenstand. Somit sind die Dokumentationseinheiten eindeutig in einem oder wenigen Worten benannt und es besteht nicht die Notwendigkeit zur Disambiguierung.

### 9.7.3 Struktur des geplanten Ordnungssystems

Das System soll eine quantitative Inhaltsanalyse der Stichproben aus beiden Enzyklopädischen Werken ermöglichen und darüber hinaus als Wertebereich für die Variable „Domain“ im Kontextmodell dienen.

Bevor die Struktur erstellt wird muss klar sein, nach welchen Kriterien die geordneten Lemmata eventuell dargestellt werden sollen. Wenn jedem Lemma nur genau ein Attribut zugeordnet wird, also es verschiedene Klassen gibt, die aber alle unverbunden und auf einer Ebene sind, dann kann auch nur nach diesem einen Attribut sortiert werden. Hier sollen die Stichproben vor allem nach zwei Attributen sortiert werden:

1. Artikelgegenstand
2. Themenbereich des Artikelgegenstand

Das Ziel des Ordnungssystems ist, die vorhandenen Begriffe zu sortieren, nicht, ein kontrolliertes Vokabular zu entwerfen. Um die Benutzung des Ordnungssystems konsistenter zu machen, ist es notwendig, zu den verwendeten Klassen Äquivalenzklassen zu bilden, deren Objekte der eigentlichen Klasse zugeordnet werden.

Alle Instanzen der Klassen schließen sich gegenseitig aus. Für jedes Lemma muss jede Kategorie ausgefüllt werden. Ist eine Person oder ein Ereignis keinem der genannten Bereiche zuzuordnen, bekommt es das Attribut „Allgemein“. Jedes Lemma wird einer Klasse vom Typ „Typ“ zugeordnet. Hier gibt es keine Auffangklasse. Bereich, Zeitraum und Nationalität sind weitere Facetten, in die das jeweilige Stichwort eingeordnet werden soll.

Strukturell gesehen ist die Haupteigenschaft eines Stichworts die Zuordnung eines Typs (etwa Person oder Gegenstand). Die Facette „Bereich“ bestimmt den Typ genau-

er, damit sind bereits die zwei wesentlichsten Punkte der Zuordnung erfasst. Die Erfassung von Zeitraum und Region fügt dann noch zusätzliche Informationen hinzu.

#### 9.7.4 Das Kategoriensystem und seine Zuweisungsregeln

Es entstand so ein System mit vier diskunkten Klassen. Jedes Stichwort wird dabei in jeder Kategorie in eine der dort vorhandenen Klassen einsortiert. Zur besseren Eindeutigkeit werden Äquivalenzklassen angegeben.

Das Lemma wird einem *Typ* zugeordnet. Dabei sind die Äquivalenzklassen zu beachten. Ein Lemma kann nur EINEM Typ zugeordnet werden.

Für jedes Lemma muss *immer* und ohne Ausnahme ein Typ angegeben werden.

Tabelle 9-2 Kategorie "Typ"

Typ	Ort [o]	Person [p]	Ereignis [e]	Ding [stofflich] [ds]	Ding [nicht stofflich] [dn]	Lebewesen [Tier, Pflanze] [l]	Stadt [s]	Personengruppe [pg]	
Äquivalenzklassen	Gewässer	Lebende und verstorbene Einzelpersonen	Einmalige und wiederkehrende Ereignisse, diskret und kontinuierlich	Gegenstände	Methoden, Entwicklungen,	Fauna und Flora	Dorf, Markt, Stadt	Mehr als eine Person	
	Region/Land		Naturereignis [incl. Naturkatastrophen]	Firmen	Filme		Siedlungen ≠Regierungsbezirk	Vereine und Verbände	
	Sehenswürdigkeit		Festivals	Stilrichtungen,	Stile, Stilrichtungen	Mikroorganismen und Bakterien		Bands und Orchester	
	Berg/Gebirge /Landschaft		Feiertage						
	Insel		Sportereignisse Zeitpunkte						

Als nächster Schritt wird die Kategorie Bereich ausgefüllt.

Tabelle 9-3 Kategorie "Bereich"

Bereich:	Äquivalenzklassen			
<b>Wirtschaft [w]</b>	Gewerbe oder Handel	Industrie	Handwerk	Geld und Währung
<b>Naturwissenschaft [n]</b>	Alle Naturwissenschaften	Medizin		
<b>Geisteswissenschaft [g]</b>	Mathematik	Sozialwissenschaft	Betriebswirtschaft	Sportwissenschaft
<b>Bildende Kunst [b]</b>	Malerei	Bildhauerei	Schmiedekunst	Töpferei
<b>Literatur [l]</b>	Journalistik	Belletristik	Poetik	Prosa
<b>Musik [m]</b>	Komponist/Kompositionen	Ausführender		
<b>Politik [p]</b>	Politikforschung	Tagespolitik		
<b>Gesellschaft/Alltag [ga]</b>	Schauspieler	Fernsehen sonstiges	Alltags	Alltagsabläufe
<b>Recht [r]</b>	Gesetze	Jura	Gerichte	Strafvollzug
<b>Technik [t]</b>	Verfahren und Maßnahmen	Techniken aller Sachgebiete	Technische Beschreibung von Gegenständen	Produktion und Instandhaltung
<b>Verkehrswesen [v]</b>	Schienenverkehr	Schifffahrt	Luftfahrt	Straßen
<b>Sport [s]</b>	Sportler	Sportereignisse	Sportvereine	Jede Art von Sport
<b>Kein bestimmter Bereich [ka]</b>	v.a. bei Lebewesen und Städten/Orten	Falls sich kein bestimmter Bereich zuordnen lässt		

Dann wird der Zeitraum ausgefüllt. Es gibt offene und geschlossene Zeitklassen. Die geschlossenen sind zu verwenden, um präzise Zeitangaben zu machen, z.B. wann ein Ereignis stattgefunden hat oder wann ein Buch erschienen ist. Für Personen und andere Phänomene von einer Lebensdauer, die länger als die möglichen auswählbaren Zeiträume sind, gilt stets der letzte aktuelle Zeitraum, in dem sie wirkten oder wirken, es kommen auch offene Klassen in Frage. Ansonsten sollten die offenen Klassen verwendet werden, wenn ein Phänomen in der Vergangenheit begann und bis heute andauert, beispielsweise: Erfindung des Rades. Alle Zeiträume sind zur besseren Verwendung kodiert. Falls sich ein Phänomen, beispielsweise ein Lebewesen („Schnecke“), keinem bestimmten Zeitraum zuordnen lässt, muss „kein bestimmter Zeitraum“ angegeben werden. Das gilt auch, wenn der Zeitraum sich aus irgendeinem Grund nicht feststellen lässt. Alle Belange *vor* 1000 v.Chr. werden mit dem Attribut „vor 1000 v. Chr.“ versehen.

Tabelle 9-4 Kategorie "Zeitraum"

Geschlossene Zeitklassen	1000 v.Chr. – 0	1	1941 – 1950	18	Offene Zeitklassen	1000 v.Chr. - heute	26
	1 n.Chr. – 200 n.Chr.	2	1951 – 1960	19		0 - heute	27
	201 – 500	3	1961 – 1970	20		250 – heute	28
	501 – 750	4	1971 – 1980	21		500 - heute	29
	751 – 1000	5	1981 – 1990	22		750 - heute	30
	1001 – 1250	6	1991 – 2000	23		1000 - heute	31
	1251 – 1500	1	2001 – 2010	24		1250 – heute	32
	1501 – 1600	2	2011 – 2020	25		1500 – heute	33
	1601 – 1700	3				1700 – heute	343
	1701 – 1750	7				1800 – heute	35
	1751 – 1800	8				1850 – heute	36
	1801 – 1820	9				1900 – heute	37
	1821 – 1840	10				1950 – heute	38
	1841 – 1860	11				1960 – heute	39
	1861 – 1880	12				1970 – heute	40
	1881 – 1900	13				1980 – heute	41
	1901 – 1910	14				1990 – heute	42
	1911 – 1920	15				2000 - heute	43
	1921 – 1930	16				Kein bestimmter Zeitraum	0
	1931 – 1940	17				Vor 1000 v.Chr.	-1

Als letzter Schritt wird die *Regionalität* zugeordnet. Es muss immer das genaueste der vorhandenen Attribute verwendet werden. Je kleiner die bezeichnete Region, desto genauer ist das Attribut, also Deutschland ist eine genauere Angabe als Europa. Das Attribut „kein bestimmtes Land“ wird angegeben, wenn das Lemma in der Tat keinem Land zugeordnet werden kann, wie zum Beispiel „Wolke“. „International“ soll verwendet werden, wenn ein Lemma *mehreren* Ländern zugeordnet werden kann.

Tabelle 9-5 Kategorie „Regionalität“

Deutschland
Österreich
Schweiz
Europa
Afrika
Amerika Nord
Amerika Süd
Australien/Ozeanien
Asien
Antarktik
Kein bestimmtes Land
international

### 9.7.5 Untersuchungen der Inhaltsverteilung mit Hilfe des Kategoriensystems

Es wurden zwei Stichproben untersucht. Diese enthielten 1000 Artikel aus der deutschsprachigen Wikipedia<sup>30</sup> (durch Verwendung der Funktion „zufälliger Artikel“) und 500<sup>31</sup> Artikel aus der Brockhaus Enzyklopädie (hier wurde ein Zufallszahlengenerator verwendet, um jeweils die Band-, Seiten- und Artikelnummer (auf der Seite) zu erhalten). Nachgeschlagen wurden in der 17. Auflage, Printversion. Dabei galt, dass nur Artikel verwendet wurden, die auf der durch den Zufallsgenerator genannten Seite vollständig vorhanden waren oder begonnen haben. Da in der Brockhaus Enzyklopädie viele Artikel über mehrere Seiten gehen, wären diese sonst bevorzugt gewesen. Die Artikelgegenstände wurden dem Kategoriensystem entsprechend annotiert (siehe Abbildung 9-3). Vor allem der Typ des Artikelgegenstands und der Themenbereich waren hier von Interesse.

<sup>30</sup> Zeitraum: Juni – August 2013.

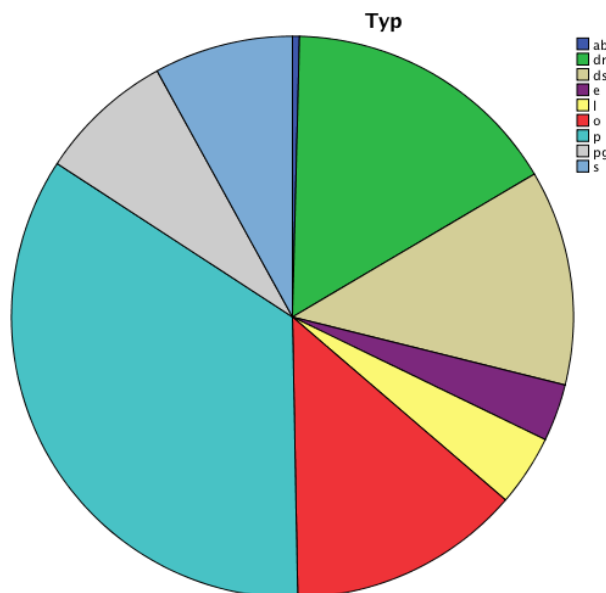
<sup>31</sup> Da die Wikipedia mittlerweile fast fünfmal so viele Artikel (Stand: Juni 2013) umfasst wie die Brockhaus Enzyklopädie wurde die Stichprobe aus der Wikipedia größer angelegt.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	c	Typ	Bereich	Zeitraum		Land		w	n	g	r	t	v	ga	i	b	s	p	m
2	Mother North	dn	m	42				0	0	0	0	0	0	0	0	0	0	0	0
3	Demokratik Toplum Partisi	pg	p	24				0	0	0	0	0	0	0	0	0	0	0	0
4	Gradmessung	dn	n	33				0	0	0	0	0	0	0	0	0	0	0	0
5	Schloss Leombach	o	k	30		d		0	0	0	0	0	0	0	0	0	0	0	0
6	Andreas Dung-Lac	p	g	9	10			0	0	1	0	0	0	0	0	0	0	0	0
7	Baise	o	k	0		f		0	0	0	0	0	0	0	0	0	0	0	0
8	Clara Eggart	p	ga	1				0	0	0	0	0	0	1	0	0	0	0	0
9	Rubidiumamid	ds	n	0				0	0	0	0	0	0	0	0	0	0	0	0

Abbildung 9-3 Kategorisierung der Artikelstichproben

## Ergebnisse

Die Stichproben wurden entsprechend der Fragestellung nach dem Gegenstand und dem thematischen Bereich, aus dem der Gegenstand stammt, ausgewertet.

Abbildung 9-4 Verteilung des Bereichs Gegenstands "Typ" eines Artikels der Wikipedia-Stichprobe<sup>32</sup>

Die Auswertung der Häufigkeiten der klassifizierten Artikel der Wikipedia-Stichprobe zeigt, dass ein Großteil der Artikelgegenstände *Personen* sind (34,5%, siehe Tabelle 9-2), wie bereits die allgemeine Betrachtung der Artikel-Kategorien in der deutschsprachigen Wikipedia vermuten ließ (siehe Kapitel 9.7.1). Am zweithäufigsten kamen *nichtstoffliche Dinge* vor (16,1%). Bei *nichtstofflichen Dingen* handelt es sich nicht um Gegenstände, sondern beispielsweise um Methoden oder Formen (z.B. „Traditionalismus“<sup>33</sup>,

<sup>32</sup> Für die Abkürzungen der Typen siehe entsprechende Tabellen im Kategoriensystem.

<sup>33</sup> Beispiele aus den Stichproben.

„Lernstrategie“, „Wasseraufbereitung im Schwimmbad“ aber auch Filme oder Musikstücke).

Mit nur 0,4% waren *Abkürzungen* in der Stichprobe aus der deutschsprachigen Wikipedia vertreten, ein Typ, der in der Stichprobe der Brockhaus Enzyklopädie nicht vorkam. Auch *Orte* waren mit einem Anteil von 13,4% häufig Gegenstand eines Wikipedia - Artikels, dabei wurden *Städte* im Sinne von Menschengesiedlungen noch einmal extra sortiert. *Orte* sind zum Beispiel Bäche und Flüsse („Kleiner Rossbach“), Straßen („Kleine Burgstrasse“), Bahnhöfe verschiedenster Art, aber auch Cafes und Restaurants. In der Stichprobe aus der BE<sup>34</sup> waren *Orte* nur mit 7,4% vertreten. Hierbei handelte es sich vor allem um Inseln, Bäder oder Landschaften.

Die restliche Verteilung auf die unterschiedlichen Typen entsprechend dem Kategoriensystem ist der Tabelle 9-6 zu entnehmen:

Tabelle 9-6 Verteilung der Artikel auf die verschiedenen Typen, WP und BE in %

Typ	Wikipedia [in %]	Brockhaus [in %]
Person (p)	34,5	21,8
Ding, nichtstofflich (dn)	16,1	30,0
Ding, stofflich (ds)	12,4	18,6
Ereignis (e)	3,3	1,2
Personengruppe (pg)	7,8	5,8
Ort(o)	13,4	7,4
Stadt (s)	8,0	8,4
Lebewesen (l)	4,1	6,8
Abkürzung (ab)	0,4	0,0
Summe	100	100

<sup>34</sup> Für Abkürzungen siehe Abkürzungsverzeichnis.

In der Stichprobe aus der Brockhaus Enzyklopädie war am häufigsten das *nichtstoffliche Ding* vertreten (30 %), gefolgt von *Personen* mit 21,8%.

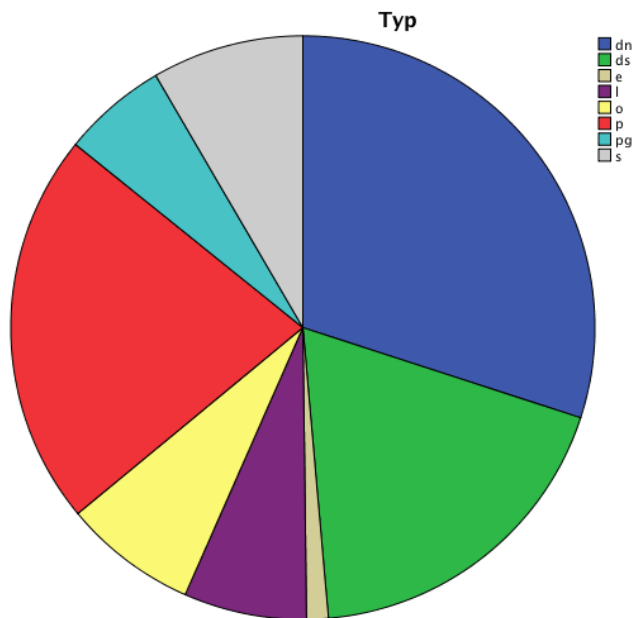


Abbildung 9-5 Verteilung des Bereichs "Typ" der Stichprobe aus der Brockhaus Enzyklopädie<sup>35</sup>

<sup>35</sup> Für Abkürzungen der Bereiche siehe entsprechende Tabelle im Kategoriensystem.



Interessant ist auch eine Betrachtung der Kombination von *Bereich* und *Typ*. Gerade bei Personen, die in beiden Werken stark vertreten sind, ist es wichtig, sich den Bereich, aus dem die Personen stammen, genauer anzusehen.

Tabelle 9-7 Verteilung der Kombination „Person“ + Themenbereiche WP und BE in %

Typ	Bereich	Wikipedia [in %]	Brockhaus Enzyklopädie [in %]
p+	Wirtschaft	3,2	0,9
p+	Naturwissen- schaft	5,6	12,8
p+	Geisteswissen- schaft	15,0	22,0
p+	Recht	1,8	1,8
p+	Technik	0,0	1,8
p+	Verkehrswesen	0,0	0,0
p+	Gesellschaft und Alltag	10,0	3,7
p+	Literatur	7,0	20,2
p+	Bildende Kunst	5,3	13,8
p+	Sport	22,3	0,0
p+	Politik	22,0	19,3
p+	Musik	7,9	3,7

Auffällig ist hier, dass bei der Wikipedia-Stichprobe 22,3% aller aufgeführten Personen aus dem Bereich *Sport* stammten (siehe Tabelle 9-6). Bei der BE dagegen sind es 0,0%. Aus dem Bereich *Literatur* wiederum waren 20,2% der Personen in der BE, in der Wikipedia lediglich 7% vertreten. Um den Fokus hier wieder auf die „Freizeitthemen“ wie *Sport* und *Musik* zu lenken: Hier ergaben sich in der Wikipedia insgesamt 30,2% aller Personen für diese zwei Bereiche zusammen und „nur“ 3,7% in der BE.

Der Chi-Quadrat Verteilungstest<sup>36</sup> für die prozentualen Vorkommen der Personen aus den Unterschiedlichen Bereichen ergibt eine Irrtumswahrscheinlichkeit dafür, dass fälschlicherweise die Annahme der Gleichverteilung der Daten bei beiden Stichproben abgelehnt wird, von  $p = 0,00E+00$ .

<sup>36</sup> Der Chi-Quadrat-Verteilungstest wird hier folgendermaßen angewendet: die Erste Datenreihe wird als „erwartete Werte“ verwendet und die beobachteten empirischen Werte der Zweiten Datenreihe werden auf Gleichverteilung gegenüber den „erwarteten“ (hier: weil in der Wikipedia so vorhandenen) Werten getestet.

Die Verteilung der Artikel auf die Themenbereiche ist ebenfalls mit einer Irrtumswahrscheinlichkeit von  $p = 1,72 \text{ E-}29$  in beiden Stichproben nicht gleichverteilt, genauso auf die der Artikeltypen,  $p = 0,001$ .

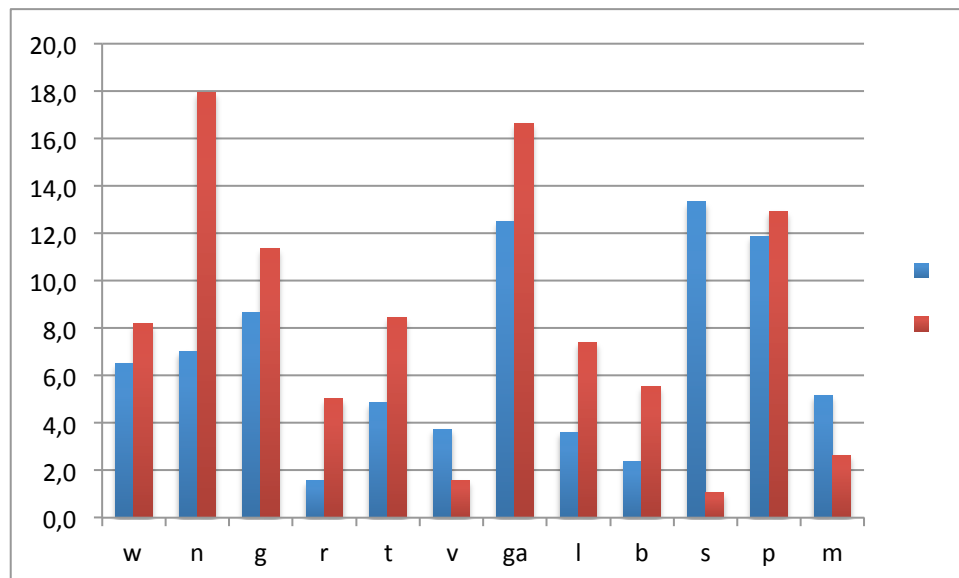


Abbildung 9-6 Verteilung der Artikel der Stichproben auf die Themenbereiche, in %, blau= WP, rot=BE

Während bei der BE der größte Anteil der Artikel aus dem Bereich *Naturwissenschaft* (z.B. „Tiergartenbiologie“, „Ruhemasse“, „Pelargonsäure“, 17,9%) stammte, gefolgt von *Gesellschaft und Alltag* (z.B. „Vermehrungsriten“, „Tradition“, „Steinkrüge<sup>37</sup>“, 16,6%), so zeigte sich bei der Klassifikation der Artikel aus der Wikipedia-Stichprobe, dass der größte Anteil (13,3%) *Sport* („Horst Funk“, „Dansk Floorball Union“, „Wimbledon Championships 1971“) zugeordnet werden konnte, dicht gefolgt von *Gesellschaft und Alltag* (12,5%) und *Politik* („Max Brandl“, „Provinzialstände (Dänemark)“, „CdZ-Gebiet Kärnten und Krain“, 11,9%). Bei den Artikeln aus dem Bereich *Politik* handelte es sich, bis auf drei Ausnahmen, ausschließlich um Personen oder Personengruppen. Bei Artikeln aus dem Bereich *Wissenschaft* zeigte sich ebenfalls eine Diskrepanz zwischen den beiden Enzyklopädien: waren in der Stichprobe der Wikipedia insgesamt 15,7% aller Artikel aus einem wissenschaftlichen Bereich, so konnten aus der BE Stichprobe 29,2% entweder den Geistes- oder Naturwissenschaften zugeordnet werden.

<sup>37</sup> In der BE wird i.d.R. der Plural verwendet.

Zu beachten ist, dass es Artikelgegenstände gibt, die keinem bestimmten Bereich zugeordnet werden können. Das waren bei der Wikipedia immerhin 18,7%, bei der BE dagegen nur 1,3%. Kein bestimmter Bereich bedeutet, dass theoretisch mehrere Bereiche vorkommen können, je nach Anwendung des Artikelgegenstands. Beispiele sind hierfür etwa Abkürzungen oder Orte (Ort als Typ ist zwar selbsterklärend, allerdings wäre es zum Beispiel bei Orten von Ereignissen möglich, sie nach dem Bereich des Ereignisses zu sortieren).

Tabelle 9-8 Zuordnung der Artikel zu den Bereichen Wikipedia und BE

Bereich	Wiki (in %)	BE (in %)
Wirtschaft	6,5	8,2
Naturwissenschaft	7,0	17,9
Geisteswissenschaft	8,7	11,3
Recht	1,5	5,0
Technik	4,9	8,4
Verkehr	3,7	1,6
Gesellschaft und Alltag	12,5	16,6
Literatur	3,6	7,4
Bildende Kunst	2,4	5,5
Sport	13,3	1,1
Politik	11,9	12,9
Musik	5,2	2,6
Gesamt	81,2	98,7
kein bestimmter Bereich	18,7	1,3
Gesamt + kein bestimmter Bereich	99,9	100,0

Allgemein zeigt die Auswertung der Stichproben, dass die inhaltliche Verteilung der beiden Informationssysteme unterschiedlich ist. Die genauere Betrachtung der Themenbereiche und Typen macht, gerade durch die starke Präsenz der Bereiche *Sport* und *Ort*, in den vor allem auch Cafes, Sehenswürdigkeiten etc. fallen, in der deutschsprachigen Wikipedia deutlich, dass Themen, die vermehrt in den Bereich „Freizeit“ fallen, in der Wikipedia stärker vertreten sind.

Für nähere Informationen – z.B. wie viele Personen aus dem Bereich *Musik* eher der Pop-Musik oder anderen modernen Strömungen zuzuordnen sind – müsste eine noch genauere Klassifikation vorgenommen werden. Dies ist dann kontextspezifischer zu machen, also konzentriert auf den entsprechenden Bereich (z.B. Musik, Sport, Literatur

etc.). Gerade bei den *Orten* geht durch die Klassifikation Information verloren, da Bereiche wie etwa „Gastronomie“ unter *Gesellschaft und Alltag* fallen. Dass diese Erkenntnisse aber relevant sind, wurde erst *nach* Durchführung und Klassifikation der Stichprobe klar und sollte bei weiteren Untersuchungen berücksichtigt werden.

Hier war das Ziel der Untersuchungen, einen allgemeinen Überblick über die inhaltliche Verteilung der Artikel in den Enzyklopädien zu schaffen. Selbst ohne Berücksichtigung weiterer Unteraufteilungen, wie oben beschrieben, bestätigte sich die Vermutung, dass sich die mengenmäßigen Schwerpunkte der Themen bei den beiden Werken anders verhalten und dass die Inhalte der Wikipedia auch zu einem großen Teil Themen abdecken, die eher einer Beschäftigung in der Freizeit zuzuordnen sind. Dieser Aspekt ist eine weitere Grundlage für die Bewertung der Annahme, dass die deutschsprachige Wikipedia sowohl für *work-based* als auch für *non-work-based* Szenarien im Sinn des ELIS [Spink und Cole, 2001] verwendet wird.

## 10 Bilder bei der Suche in der Wikipedia

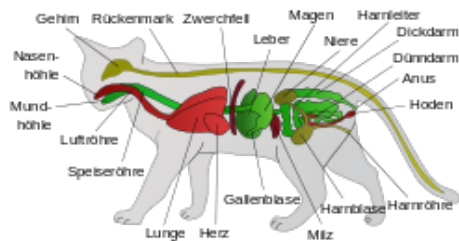


Abbildung 10-1 Anatomie einer Katze, Uwe Gille, Quelle:

[http://upload.wikimedia.org/wikipedia/commons/8/8b/Scheme\\_cat\\_anatomy-de.svg](http://upload.wikimedia.org/wikipedia/commons/8/8b/Scheme_cat_anatomy-de.svg)



Abbildung 10-2 schlafende Babykatze, Quelle:

[http://upload.wikimedia.org/wikipedia/commons/f/fb/Sleeping\\_baby\\_cat.jpg](http://upload.wikimedia.org/wikipedia/commons/f/fb/Sleeping_baby_cat.jpg)

Ein zentraler Aspekt dieser Arbeit beschäftigt sich mit der Frage nach den verschiedenen Präferenzen der Nutzer für Inhaltselemente in der Wikipedia abhängig vom Nutzungsszenario. Vor allem beim Medientyp Bild wird ein Unterschied in der Nutzung vermutet, denn gerade Bilder haben sehr viel Potential, abzulenken, aber auch zu ergänzen und zu erfreuen. Sie können stark oder weniger stark emotional sein – und haben auch das Potential, Emotionen hervorzurufen [Kappas und Müller, 2006]. Da in dieser Arbeit Emotionen sowohl als Einfluss als auch als Zielgröße relevant sind, ist eine nähere Betrachtung des in diesem potentiell einflussreichen [Kappas und Müller, 2006] Inhaltelements Bild sinnvoll. Eine Untersuchung von Schubart und Knäusl [Knäusl und Schubart, 2013] zeigte dazu, dass die Qualität des Bilderangebots während der Suche (hier bei einheitlichen Suchsituationen, in etwa entsprechend dem Typ

*learn* nach [Marchionini, 2006]) deutlich mit der emotionalen Selbstauskunft der Nutzer korrelierte [Knäusl und Schubart, 2013]. Wie die zwei Katzenbilder (siehe Abbildung 10-1 und Abbildung 10-2) veranschaulichen, haben Bilder ein emotional unterschiedliches Potential, obwohl sie einen semantisch ähnlichen oder gleichen Gegenstand abbilden. Schubart [Schubart, 2013] zeigte dazu, dass die Emotionalität, die ein Bild auslöst, einen Einfluss auf die subjektive Bewertung eines Bildes durch den Nutzer hat.

Für die vom Nutzungsszenario abhängige Eignung von Bildern sind – so die Hypothese hier – zwei wesentliche Punkte verantwortlich:

1. Die generelle und die situationsgebundene Funktion eines Bildes
2. Bildinterne Faktoren, die die subjektive Bewertung des Nutzers beeinflussen.

Dies ist wiederum ebenfalls abhängig von dessen kognitiver Struktur.

[Knäusl und Schubart, 2013] zeigten zum Beispiel, dass in gegebenen Szenarien fast doppelt so oft dekorative oder veranschaulichende (Funktion) Bilder betrachtet wurden als Bilder, die zusätzliche, nicht im Text vorhandene Informationen enthielten (ergänzende Funktion). Auch wenn hier die Gründe dafür nicht direkt ersichtlich werden, so ist doch eine Berücksichtigung von Bildern bei der Suche in der Wikipedia abhängig von deren Funktion zu erkennen.

Ein weiterer Grund für die differenzierte Betrachtung von Bildern ist die Tatsache, dass zumindest eine Fotografie formal immer nur eine Instanz abbildet [Nöth, 2008], allerdings eine Klasse repräsentiert. Ein Bild von einer Katze enthält nur *eine* Instanz der Klasse Katze, hinzukommt, dass diese Instanz wiederum stark unterschiedlich dargestellt werden kann. Hier liegt ein Unterschied zum Text, der zwar ebenfalls unterschiedlich verfasst werden kann, allerdings im Bereich der Online-Enzyklopädie stark sachlich ist und damit im Gegensatz zu den Bildern in dieser Frage nicht differenzierter betrachtet wird.

Da die Unterschiede in den Nutzerpräferenzen deshalb vor allem beim Medium Bild vermutet werden und auch hier davon ausgegangen wird, dass die Relevanz nicht binär ist und von mehreren (als nur dem reinen Bildgegenstand) Faktoren abhängig ist,

wird die Bildauswahl zum einen und die möglichen Funktionen von Bildern zum anderen hier besprochen.

## 10.1 Bilder in der Wikipedia

Gerade die Bebilderung der Artikel ist in der Wikipedia ein großer Unterschied zu traditionellen, gedruckten Enzyklopädien. Viele Artikel sind nicht nur umfangreich bebildert, sondern unter Umständen sogar überfrachtet. Das kann den Umfang eines Artikels erhöhen, ohne dass wesentlich zum Inhalt beigetragen wird. Dieses „Problem“ kann ebenfalls dazu beitragen, dass ein Artikel an Übersichtlichkeit verliert und dem Nutzer die Selektion der situativ Relevanten Elemente erschwert wird.

Die Bilder, die in den Artikeln der Wikipedia verwendet werden, entstammen der Bildersammlung der *Wikimedia Commons*<sup>38</sup>. Dabei handelt es sich um eine Sammlung von Mediendateien, zu der jeder beitragen kann. Es gibt klare Richtlinien, welche Dateien in den *Commons* aufgenommen werden, sie sollen in erster Linie edukativer Natur sein und zweitens nicht „besser in einem anderen Projekt der Wikimedia-Foundation aufgehoben sein“<sup>39</sup>.

Grundsätzlich darf jeder Autor nur sein *eigenes* Werk dort hochladen und er stellt es dann unter eine freie Lizenz, sprich es darf unter bestimmten Bedingungen (diese werden z.B. durch die *Creative Commons Licences*<sup>40</sup> festgelegt) weitergegeben und verwendet werden. Es werden nur gemeinfreie oder gewissen Einschränkungen, wie z.B. Namensnennung des Urhebers bei Wiederverwendung, unterworfenen Werke dort akzeptiert.

Die Auswahl der Bilder für die Artikel in der Wikipedia erfolgt nach den Wikipedia-internen Kriterien. Es wird betont, dass ein gutes Verhältnis zwischen Text und Bild eingehalten werden soll, wie dieses allerdings aussieht, bleibt offen.

Ein Bild kann lt. Richtlinien dann im Artikel verwendet werden, wenn es einen inhaltlichen Bezug zum Artikelgegenstand hat, der Bildinhalt relevant für den Text ist und zum besseren Verständnis beiträgt, wenn es Merkmale des Artikelgegenstands abbil-

---

<sup>38</sup> Wikimedia Commons [[http://commons.wikimedia.org/wiki/Main\\_Page](http://commons.wikimedia.org/wiki/Main_Page)], (abgerufen 16.Mai 2013).

<sup>39</sup> Projektrahmen Wikimedia Commons [<http://commons.wikimedia.org/wiki/Commons:Projektrahmen>], (abgerufen 16.Mai 2013).

<sup>40</sup> Für weitere Informationen siehe: [<http://creativecommons.org>], (abgerufen 16.Juli 2012).

det, die für diesen wesentlich und identitätsstiftend sind, wenn das Motiv des Bildes auch tatsächlich das im Text angesprochene ist (Authentizität des Motivs) und wenn das Bild sich gegenüber allen anderen bereits im Artikel vorhandenen Bildern unterscheidet, also nicht bereits ein ähnliches oder vergleichbares Bild im Artikel vorhanden ist<sup>41</sup>. Diese Regeln bilden die allgemeine Grundlage für die Auswahl und Einbindung von Bildern in einen Artikel. In diesem Kapitel wird später gezeigt, dass für den Nutzer subjektiv weitere Auswahlkriterien relevant sind.

Es wird in den Richtlinien weiterhin ausdrücklich darauf hingewiesen, dass es sinnvoller ist, auf die *Wikimedia Commons* zu verweisen, als zu viele Bilder in einen Artikel aufzunehmen.

Die Bilder in den *Commons* sind nicht zwangsweise auch von demjenigen, der sie verwendet, erstellt worden. Es gibt eine große Menge an Bildern, von denen nicht klar ist, wer ihr Autor ist. Diese Bilder kommen meistens aus *public domain* Quellen und werden zum Beispiel auf den Websites öffentlicher Einrichtungen wie Regierungen oder Hochschulen frei verfügbar angeboten. Auch gemeinfreie Bilder aus Büchern oder Zeitschriften werden von Usern digitalisiert und in den *Commons* zur Verfügung gestellt [Viegas, 2006]. Es gibt deshalb auch sehr gute Bilder in der Wikipedia, oftmals für einen kleinen Bereich überdurchschnittlich viele, sehr gute Bilder [Viegas, 2006]. Diese stammen von Bild-Autoren, die tatsächlich wissenschaftlich wertvolle und hochwertige Bilder der *Wikimedia Foundation* zum Nutzen der allgemeinen Öffentlichkeit spenden. Eine große Vielfalt aber auch eine schwankende Qualität der Bilder ist damit gegeben.

Die Richtlinien und Hinweise zur Verwendung von Bildern in den Wikipedia-Artikeln sind eigentlich relativ gut. Demnach könnte man davon ausgehen, dass die Bebilderung der Artikel stimmig sein müsste. Die Realität zeigt allerdings, dass dies nicht immer der Fall ist. Autoren verwenden gerne ihre eigenen Bilder in Artikeln, auch wenn diese nicht unbedingt den gewünschten Kriterien entsprechen. Deshalb gibt es Artikel, in denen eine große Zahl von Bildern vorhanden ist, ohne dass diese nennenswerten Mehrwert zu haben scheinen. In diese Richtung gehen auch die Ergebnisse von

---

<sup>41</sup> Wikipedia: Bilder [<http://de.wikipedia.org/wiki/Hilfe:Bilder>], (abgerufen 15. November 2013).



[Knäusl und Schubart, 2013], die zeigten, dass jeweils weniger oder weit weniger als die Hälfte aller vorhandenen Bilder überhaupt in die Erledigung des Suchtasks einbezogen wurden. Begründungen dafür liegen allerdings nicht vor, weshalb diese hier weiter untersucht wurden (siehe ab Kapitel 10.8).

## 10.2 Der Begriff „Bild“

*„Unter einem Bild verstehen wir nichtsprachliche Formen der visuellen Kommunikation....“*  
[Ballstaedt, 2005, S.14].

Für die folgenden Arbeiten ist es für ein einheitliches Verständnis notwendig, eine kurze Definition des Bildbegriffs zu geben. Viele Disziplinen befassen sich mit dem Thema Bild, so dass die Definitionen durchaus unterschiedlich ausfallen. Deshalb ist es sinnvoll, eine möglichst allgemeine, aber restriktive Beschreibung zu finden, die aber weitestgehend auf philosophische Aspekte verzichtet. Die Untersuchungen beschränken sich auf digitale Bilder.

Über die digitalen Bilder sagt Heidenreich in Sachs-Hombach, [Heidenreich, 2005 S. 382]:

*„Digitale Kameras, Mobiltelefone etc. beruhen alle auf derselben Verknüpfung von binären Prozessoperationen, Bitcode und einer Oberfläche von Sichtbarkeit. Zwischen Bild-, Text- und Tondaten bestehen an der Basis digitaler Verarbeitung keine Unterschiede mehr“.*

Digitale Bilder liegen kodiert in Bitcode vor und werden durch entsprechende Interpretation dargestellt. Damit sind in die Definition der digitalen Bilder hier auch digitale Abbilder analoger Bilder eingeschlossen. Eine inhaltliche Unterscheidung (nach dem Entstehungsprozess) findet in der Klassifikation zur Verortung von Bildern (siehe Anhang 1) statt.

Bilder haben nach Sachs-Hombach weiterhin die wichtigen Eigenschaften, künstlich hergestellt, flächig und dauerhaft zu sein [Sachs-Hombach, 2001]. Dauerhaft bedeutet in diesem Fall vor allem, dass sie wiederholt aufgerufen werden können. Ein Merkmal von Bildern muss darüber hinaus sein, dass es Bildelemente enthält. Diese Elemente sind der Gegenstand des Bildes, anhand dessen es zum Beispiel klassifiziert bzw. beschrieben werden kann.

### 10.3 Verwendung von Bildern

Bilder spielen in vielen Disziplinen eine Rolle. Auch in der Informationswissenschaft und im Rahmen des Information Retrieval bzw. Multimedia Retrieval sind Bilder Gegenstand der aktuellen Forschung (z.B. [Cai et al., 2012]).

Das Nebeneinander von Text und Bild ist unter pädagogischen Aspekten bzw. im Zusammenhang mit Lerntheorien mitunter ein entscheidender Faktor. Nöth [Nöth, 2008] beschreibt dieses Zusammenspiel von Text und Bild als „...den Beitrag des einen für das andere Medium im Rahmen einer Gesamtbotschaft“ [Nöth, 2008, S.492].

Traditionelle enzyklopädische Texte wie in der Brockhaus Enzyklopädie sind tendenziell nicht so stark bebildert, da jedes Bild Kosten in Anschaffung und Druck verursacht und darüber hinaus einigen Platz beansprucht. Generell gilt aber, dass die traditionell knappe Formulierungsweise enzyklopädischer Texte von einer Anreicherung mit Bildmaterial profitiert.

In den digitalen Medien dagegen sind Bilder – abgesehen von Urheber- und Verwertungsrechtsfragen und der entsprechenden generellen Verfügbarkeit – mühelos bunt und großzügig zu verteilen, ohne dass erhebliche Mehrkosten oder Platzprobleme auftreten. Dabei ist deshalb manchmal nicht mehr klar, ob ein Bild in einem digitalen Medium wie z.B. einer Website oder einem Wikipedia-Artikel nur deshalb dort platziert wurde, weil es eben vorhanden und/oder zudem hübsch war, oder ob es wirklich zur Verdeutlichung eines Sachverhaltes beiträgt.

Nicht einmal die Richtigkeit eines Bildes, im Sinne von „der abgebildete Gegenstand entspricht dem Artikel/Seiten-Inhalt und/oder gehört zum Kontext des Inhaltes“ bzw. „enthält neue, noch nicht an anderer Stelle vorhandene Information“ kann man allzu leicht feststellen und noch viel weniger voraussetzen.

Darüber hinaus besteht die Vermutung, dass auch syntaktische Faktoren Einfluss auf die Eignung oder den Nutzen eines Bildes haben können. Ein syntaktischer Faktor ist etwa die Größe eines Bildes oder die Frage, ob es sich um ein Farbbild oder ein schwarz-weißes Bild handelt. Auf pragmatischer Ebene werden Fragen nach der Bildverwendung untersucht, also zum Beispiel, ob das Bild nur eines von vielen ist oder einzeln an prominenter Stelle platziert wurde (siehe auch [Nöth, 2008]). Unter Umständen kann ein an sich gutes Bild zum Beispiel in der Reihe vieler Bilder eher zur Verwirrung des Betrachters beitragen als zur Verbesserung der Verständlichkeit und/oder der Behaltensleistung.

Georg Peez beschreibt [Peez, 2006, S. 18]:

*„die Analyse des visuellen Materials wird mit der Analyse verbalsprachlichen Materials verifiziert. Bild und Textanalyse sind demnach nicht als konkurrierende, sondern als ergänzende Interpretationswege zu verstehen.“*

Der ergänzende Aspekt von Bildern ist wichtig, da nur dieser es möglich macht, individuelle Entscheidungen über die Verwendung von Bildern zu treffen. Diese subjektive Bewertung von Bildern als ergänzendes Moment bei der Informationsrecherche ist hier von Interesse.

Nach [Peez, 2006] gilt für die Analyse von Bildern bzw. für deren Interpretation das Verfahrensprinzip der Objektiven Hermeneutik. Entscheidend dafür ist, dass es bei der Interpretation nicht um Perspektivenübernahme und Fremdverstehen geht – ganz betont nicht um das Handeln nach bestimmten Regeln oder darum, latente Sinnstrukturen bei der Interpretation zu erkennen. Objektivität steht also nicht im Gegensatz zur subjektiven und individuellen Bewertung und Interpretation durch eine Person son-

dern bedeutet, frei von einer vorgegebenen Strukturiertheit. Auch wenn sich ein Bild optisch nicht ändert, so kann sich aber unsere Perzeption von ihm durch neue Informationen innerhalb kürzester Zeit drastisch verändern (dies auch entsprechend [Brookes, 1980; Ingwersen und Järvelin, 2005; Schamber et al. 1990; Ingwersen 1982]).

Die Vorstellung von einem Bild beruht also auch auf unseren Vorstellungen und Erwartungen, die wir an das Bild herantragen [Haasebrook, 1995]. Diese Tatsache entspricht dem Bezug des *cognitive viewpoint* [Ingwersen und Järvelin, 2005] auf das Bild. Die kognitive Struktur hat ebenfalls einen Einfluss auf den situativen Mehrwert eines Bildes (z.B. ist der Bildgegenstand bekannt, erkennbar, ist er emotional konnotiert usw.). Bilder lassen eine Mehrdeutigkeit zu, die Spielraum für Interpretationen bietet [Schnotz, 2006].

Das referiert auch auf die kontextabhängige Relevanz von Bildern, allerdings mit der Erweiterung, dass dabei nicht nur das Medium, sondern vor allem auch die Unterscheidung innerhalb des Mediums Bild nach bestimmten Kriterien eine Rolle spielt.

Im Allgemeinen sollten Bilder und Text zwar in einer synergetischen Beziehung stehen – das ist aber eben nicht immer zwangsläufig der Fall. Formale und inhaltliche Aspekte der Text-Bild-Beziehung müssen aufeinander abgestimmt werden. Oft wird zuerst das Bild erfasst, dann erst der Text [Weidenmann, 1991]. Hier entscheidet der Nutzer, abhängig von seinen Bedürfnissen. Allerdings kann ein Bild auch in einer fehlerhaften Text-Bild-Beziehung stehen, also ohne eine für die Lernförderlichkeit notwendigen Funktion (z.B. durch „falschen“ Inhalt, der nicht dem erwarteten bzw. dem der synergetischen Beziehung entsprechenden Sachverhalt darstellt). In dieser Arbeit wird dieser der Aspekt der Lernförderlichkeit noch durch die situative Angemessenheit bezüglich des individuellen Nutzerbedürfnisses erweitert. Das bedeutet, dass auch wenn die Text-Bild-Beziehung formal korrekt ist ein Bild unter Umständen situativ keinen Beitrag zum intendierten Lernerfolg liefert (siehe dazu vor allem auch Kapitel 11.3). Dazu gehört der Logik des *cognitive viewpoint* (und nach [Weidenmann, 1991]) folgend auch die aktuelle Wissensstruktur des Nutzers.

*„Ein Bild muss vom Betrachter umfassend wahrgenommen und richtig interpretiert und mit dem vorhandenen Wissen verknüpft werden“*

[Weidenmann 1991, S. 24]

Das bestärkt, dass die Unterscheidung bei Bildern nicht ausschließlich auf die Frage *ob*, sondern auch darauf, *welches* Bild verwendet wird, gerichtet ist.

#### 10.4 Individueller, situationsabhängiger Nutzen von Bildern

Im Gesamtkontext dieser Arbeit spielt es eine Rolle, wie der Nutzerkontext die Präferenz für Bilder beeinflusst, basierend auf der Aussage, dass Bilder nur lernförderlich sind, wenn sie eine bestimmte Funktion erfüllen [Levin, 1981].

Die Verwendung von Bildern (im Sinne der Berücksichtigung bei einer Suchaufgabe oder der expliziten Bewertung) wird hier in verschiedenen Studien untersucht. Dazu sind Überlegungen notwendig, z.B. zur Klassifikation von Bildern und zu Faktoren, die möglicherweise über die den allgemeinen Lerntheorien (z.B. [Paivio, 1971]) zu entnehmenden relevanten Aspekte hinaus einen Einfluss auf die Verwendung von Bildern haben. Generell geht es darum, zu bestärken, dass die Auswahl von Inhalten auch innerhalb eines Modus situations- und nutzerrelativ ist (Siehe dazu auch [Knäusl et al., 2012; Knäusl und Schubart, 2013]).

Die entscheidenden Punkte der genannten Theorien und Überlegungen sind im Zusammenhang mit dieser Arbeit:

- a) die unterschiedliche Rezeption und kognitive Verarbeitung von Bildern und Text, abhängig von der Vorstellung und Wissensstruktur des Nutzers
- b) die gegenseitige Ergänzung von Text und Bild
- c) die mögliche lernförderliche aber situationsabhängige Wirkung von Bildern
- d) die Annahme der objektiven und damit freien aber dennoch situativ und von der Wissensstruktur des Individuums geprägten Interpretation von Bildern

## 10.5 Situative Relevanz von Bildern

Im Rahmen eines Seminars im WS 2012/2013 unter der Leitung der Autorin und Professor Dr. Bernd Ludwig zeigte eine Gruppe von Studierenden, dass die Verwendung von Bildern (im Sinne des Verhältnisses von vorhandenen Bildern im Sichtfeld des Suchenden und tatsächlich betrachteten Bildern) in den verschiedenen Suchsituationen unterschiedlich ist. So wurden bei den *casual-leisure* Aufgaben<sup>42</sup> 41,3% aller vorhandenen visuellen Elemente betrachtet, bei den *look up* Tasks dagegen nur 19,8%. Bei den *learn* Aufgaben schenkten die Probanden 38,9% aller vorhandenen visuellen Elemente ihre Aufmerksamkeit [Frank et al., 2013].

Am interessantesten ist allerdings der Aspekt, dass die Varianz des Verhältnisses von vorhandenen zu betrachteten Bildern beim sog. „ziellosten browsen“ (*casual-leisure*) relativ groß ist bzw. eine Standardabweichung von 8,4 bei im Mittel 9,0 betrachteten Bildern. Das ist als Hinweis zu sehen, dass außer dem Suchtyp offensichtlich noch weitere Faktoren den Nutzer in seiner Entscheidung, wie viele der vorhandenen Bilder er berücksichtigt, beeinflussen. Als wesentliche Einflussfaktoren werden deshalb weiterhin vermutet: Erstens die Art und Funktion des Bildes und zweitens die *subjektiven* Auswahlbegründungen, warum ein Nutzer ein Bild berücksichtigt/präferiert bzw. es lernförderlich ist.

## 10.6 Gezielte Suche und Klassifikation von Bildern

Postuliert man den Nutzen einer situationsgebundenen Auswahl von Bildern, muss geklärt sein, ob und wenn ja, wie Bilder automatisch klassifiziert und ausgewählt werden können.

Hat man Erkenntnisse über die jeweiligen Eigenschaften von Bildern und deren Eignung in bestimmten Situationen, steht man weiter vor dem Problem, wie man erstens diese Eigenschaften zusammenfassen und formalisieren kann und zweitens vor der Frage, ob und wenn ja, wie eine automatische oder zumindest halbautomatische Auswahl der Bilder anhand der erhobenen Kriterien möglich ist. Für eine Umsetzung einer gezielten Bildauswahl ist das Voraussetzung.

---

<sup>42</sup> Für genauere Informationen zu Methodik, Aufbau und Durchführung der Studie siehe ab Kapitel 12.

Um eine automatische Suche nach individuellen Kriterien zu ermöglichen, müssen unter Umständen mehr Details über ein Bild bekannt sein als nur die mitgelieferten Metadaten (siehe auch [Zhang et al., 2011]).

Die im Folgenden beschriebenen Ansätze zeigen, dass es bereits einige Möglichkeiten gibt, Bilder automatisch auszuwählen. Weitere Faktoren (wie z.B. Emotionalität oder Qualität eines Bildes) müssten aber mit dem Bild vorliegen, was bei geschlossenen Systemen, aber auch bei den Inhalten der *Wikimedia Commons* vorstellbar wäre, gäbe es dafür einheitliche Verfahren.

Einige Informationen über ein Bild werden in den digitalen Medien bereits mittransportiert [Heidenreich, 2005] wie z.B. Inhalt, Herkunft, Verwendung. Durch die enge Verflechtung mit den das Bild umgebenden Daten können auch Hinweise wie Autor, Ort, Zeit und Motiv des Bildes [ebd.] oft relativ leicht automatisch einem Bild zugeordnet werden.

Formal kann man die *Eignung* eines Bildes bzw. sein Verhältnis zu einem Text zum Beispiel beschreiben wie Zhang et al. [Zhang et al., 2011]. Mit ihrem RCT-Modell haben die Autoren eine Methode entwickelt, die anhand der drei Indikatoren R (Relateness), C (Consistency) und T (Typically) die wesentliche Faktoren zur Text-Bild-Beziehung aufgreifen. Anhand einer Formel wird so der Grad der Beziehung zwischen Bild und Artikel berechnet.

Smith und Chang [Smith und Chang, 1997] haben das System VisuelSEEK zur inhaltsbasierten Bildersuche geschaffen. Der Nutzer kann eine Suchanfrage mit Hilfe inhaltlicher Kriterien formulieren. Durch einen Algorithmus werden die Inhalte der Bilder (ermittelt durch bestimmte, symbolisch gespeicherte Farbverläufe und eine Zerlegung der Bildinhalte in Teilgebiete etc.) mit der Suchanfrage abgeglichen. Zu betonen ist hier der Versuch, eine inhaltliche Suche mit einer automatischen Bildzerlegung (z.B. Welche Komponenten hat das Bild eines Löwen?) und Indexierung zu verknüpfen.

Smith und Chang [Smith und Chang, 1997] sortieren weiterhin Bilder automatisch in drei Stufen nach Klassen wie „*color photos, graphics (...)*“ (first stage: Type) „*silhouettes, center-surround (...)*“ (second stage: composition) und als drittes semantische Klassen wie „*beaches, building, nature, sunsets (...)*“. Die Informationen, um die Bilder in die korrekten Klassen einordnen zu können, werden dem dazugehörigen Text und dem Bild selbst entnommen [Smith und Chang, 1997]. Es wären hier allerdings weitere Unterscheidungen und Klassen denkbar: *Man-made, machine-made, scientific images* (z.B.: Röntgenbilder) [Müller, 2003]. In der im Rahmen dieser Arbeit und der Bachelorarbeit von Schubart [Schubart, 2012] erstellten Klassifikation zur Sortierung von Bilder (siehe Anhang 1) wird dies berücksichtigt, um eine möglichst genaue Beschreibung von Bildern zu erhalten.

WebSeer [Athitsos et al., 1996] ist ein System, mit dessen Hilfe der Versuch gestartet wurde, Bilder im Web zu beschreiben bzw. nach bestimmten bildbezogenen Kriterien aufzufinden. Neben den üblichen, teilweise ja auch nur spärlich vorhandenen Metadaten, die ein Bild an sich nur unter Umständen mittransportiert, werden sowohl der Bildinhalt als auch der mit dem Bild verbundene Text verwendet, um eine möglichst eindeutige Beschreibung des Bildes zu generieren und so ein Bild zu finden, das am besten zu den jeweiligen Nutzerbedürfnissen passt. Dies ist wieder im Sinne von [Nöth, 2008].

Einteilungen und Klassifikationen von Bildern finden sich zum Beispiel bei Heidenreich [Heidenreich, 2005], der digitale Bilder z.B. (sinngemäß) nach berechneten versus aufgenommenen oder nach semantisch engen vs. nur lose definierten Bildern unterscheidet, aber selbst feststellt, dass in der Praxis diese Typen selten in Reinform auftreten.

[Mueller, 2003] bietet eine allgemeinere Definition, in dem sie auf der einen Ebene zwischen materiellen und immateriellen Bildern unterscheidet und in einer zweiten Dimension die Bilder nach der Gestalt ihres Inhaltes, der Produktionsart und der Rezeption klassifiziert. Diese Klassifikation beschreibt mit der Rezeption einen wichtigen Aspekt, der in den anderen genannten Vorgehensweisen nicht oder nur sehr wenig berücksichtigt wird.



Es existieren einige weitere Ansätze zur Klassifikation von Bildern, die sich aber in der größeren Zahl vor allem mit dem Typus Bild *an sich*, also der Frage nach Herstellung, Intention oder Verwendung richten (Siehe dazu [Schmauks, 2005] oder [Simons, 1999]). In der Klassifikation von Knäusl und [Schubart, 2012] sollen darüber hinaus noch Faktoren bezüglich der Verwendung im Artikel und Aspekten wie der Emotionalität aufgenommen werden. Die vorgestellten Ansätze greifen alle den Gedanken der unterschiedlichen Eignung von Bildern insofern auf, als dass sie versuchen mit möglichst genauer Beschreibung und Erfassung der Bilder ein optimales Verhältnis zur Suche herzustellen.

## 10.7 Funktionen von Bildern

Neben den *Eigenschaften*, die ein Bild haben kann ist darüber hinaus seine *Funktion* für die Eignung entscheidend (vor allem für die Lernförderlichkeit [Levin, 1981]). Dass Bilder abhängig von ihrer Funktion bei der Suche in der Wikipedia unterschiedlich stark berücksichtigt werden zeigen [Knäusl und Schubart, 2013].

Funktionen von Bildern sind nach [Levin, 1981]:

- a) Verständnisfunktion (Bild hilft, den Text besser zu verstehen)
- b) Aufmerksamkeitsfunktion (Bild lenkt die Aufmerksamkeit auf bestimmte Aspekte im Text)
- c) Darstellende Funktion (Bild visualisiert und konkretisiert Inhalte des Textes)
- d) Behaltensfunktion (Bild unterstützt das Behalten von Textinformationen)
- e) Organisationsfunktion (das Bild verdeutlicht Zusammenhänge des Textes)
- f) Interpretationsfunktion (das Bild trägt dazu bei, schwierige Textabschnitte verständlicher zu machen)
- g) dekorative Funktion (das Bild dient nur zu dekorativen Zwecken)

Bilder haben ein anderes semiotisches Potential als Texte, um Sachverhalte, Gedanken, Ideen und Zusammenhänge dazustellen. So können sie gut geeignet sein, räumlich-visuelle Sachverhalte darzustellen (beispielsweise Landkarten sind schlecht textuell zu repräsentieren), allerdings repräsentieren sie im wesentlichen Visuelles, während da-

gegen Sprache gut auch andere, etwa taktile oder akustische Sinneswahrnehmungen beschreiben kann. Außerdem ist es schwierig, Bilder aufeinander zu beziehen, kausale und logische Verbindungen können nicht repräsentiert werden, dazu ist in der Regel Text notwendig [Nöth, 2008].

Ein Bild ist bis zu einem gewissen Grad pragmatisch offen, hierzu ein Zitat von [Wittgenstein, 1953, §22]:

*„Denken wir uns ein Bild, einen Boxer in bestimmter Kampfstellung darstellend. Dieses Bild kann nun dazu gebraucht werden, um jemanden mitzuteilen, wie er stehen, sich halten soll; oder, wie ein bestimmter Mann dort und dort gestanden hat; oder etc.etc. [...]“*

Diese Aussage spiegelt einen zentralen Aspekt bei der Frage, warum ein Bild ausgewählt wird, wider. Welche Inhalte sieht derjenige primär in dem Bild? Die semantische Offenheit ist eine Spezialität von Bildern. Die Funktion eines Bildes kann sich mit der jeweiligen Rezeption verändern.

In einer Studie soll deshalb untersucht werden, wie passend die in diesem Fall zur Illustration ausgewählten Bilder vom Nutzer empfunden werden, bzw. welche Funktionen die Nutzer als entscheidende für die Auswahl des Bildes zur Illustration eines bestimmten Textes angeben. Diese Untersuchung soll Aufschluss über die Gründe liefern, warum Nutzer Bilder zur Illustration von Texten auswählen. Die Vermutung, dass es unterschiedliche Kriterien für die Nutzer gibt, basiert auf dem individuellen Bildverständnis und den jeweils für die Auslegung geltenden Bedingungen.

## 10.8 Studie zur Verwendung und Auswahl von Bildern

Um die Frage nach der Funktion von Bildern bzw. der vom Nutzer präferierten Funktionen bei der Suche in der Wikipedia zu klären wurde ergänzend zur Literatur eine Studie durchgeführt. Es wurde damit auch noch einmal der Aspekt aufgegriffen, dass Bilder auch bei objektiv erfasst gleichen Kriterien (betreffend den Bildgegenstand) unterschiedlich rezipiert werden (können) und die Begründungen dafür verschieden ausfallen können. Die Untersuchungen waren wiederum konkret auf die deutschsprachige Wikipedia bezogen.

### 10.8.1 Aufbau

In der empirischen Untersuchung zur Auswahl von Bildern zur Illustration eines Wikipedia-Artikels sollten die Testpersonen zunächst vier Bilder mit den Schulnoten 1 - 6 bewerten. Die präsentierten Bilder wurden ausgewählt, indem aus dem zu untersuchenden Textabschnitt (verschiedene Abschnitte aus dem Wikipedia-Artikel über Wolfgang Amadeus Mozart<sup>43</sup>) manuell für den jeweiligen Abschnitt wichtige Schlüsselwörter herausgesucht wurden. Zu diesen Schlüsselwörtern wurden mit Hilfe der Google-Bildersuche Bilder gefunden.



Abbildung 10-3 Bild aus der Studie: Mozart and Linley 1770, Quelle:  
[http://de.wikipedia.org/wiki/Mozart#mediaviewer/Datei:Mozart\\_and\\_Linley\\_1770.jpg](http://de.wikipedia.org/wiki/Mozart#mediaviewer/Datei:Mozart_and_Linley_1770.jpg)

Die Auswahl von jeweils vier Testbilder erfolge nach zweierlei Gesichtspunkten: Erstens sollten die Bilder mehrere Aspekte (Schlüsselwörter) des Testabschnitts abdecken und zweitens verschiedene Varianten ein- und desselben Aspekts darstellen. Eines der vier Testbilder war jeweils das im Wikipedia-Artikel an der entsprechenden Stelle verwendete Bild. Ein Erkenntnisziel dieser Studie war, ob die generelle Bewertung des Bildes für die spätere Auswahl entscheidend ist (Bild gefällt oder gefällt nicht) oder ob

---

<sup>43</sup> Wikipedia: Wolfgang Amadeus Mozart [[http://de.wikipedia.org/wiki/Wolfgang\\_Amadeus\\_Mozart](http://de.wikipedia.org/wiki/Wolfgang_Amadeus_Mozart)] (abgerufen Dezember 2011).

textbezogene Kriterien (Bild ergänzt Text etc.) dominieren. Deshalb wurden die Bilder zunächst von den Nutzern *ohne* den zugehörigen Text bewertet.

Im nächsten Schritt wurde den Testpersonen der jeweilige Abschnitt aus dem Wikipedia-Artikel vorgelegt, zu dem eines der vier zuvor unabhängig beurteilten Bilder ausgesucht werden sollte. Es bestand auch die Möglichkeit, keines der vorgeschlagenen Bilder auszuwählen. Bei der Begründung für die Bildauswahl gab es neun vorgegebene Argumente, aus denen die Testperson auswählen konnte (Mehrfachnennungen waren möglich).

An der Studie (Januar-Februar 2012) haben insgesamt 16 Testpersonen zwischen 23 und 39 Jahren teilgenommen, wobei ein Fragebogen nicht verwendet werden konnte. Es wurden also insgesamt 15 x 5 Textabschnitte mit jeweils vier möglichen Bildern untersucht.

### 10.8.2 Auswertung

Die folgenden Tabellen zeigen, wie die Bilder bewertet und welche Bilder ausgewählt wurden, um die Textabschnitte zu illustrieren.

Allgemeine Ergebnisse:

Bildauswahl	Bild 1	Bild 2	Bild 3	Bild 4	Kein Bild
1. Abschnitt	<b>10</b>	1	2	2*	
2. Abschnitt	<b>5</b>	1	<b>5</b>	3*	1
3. Abschnitt	0	2	<b>13*</b>	0	
4. Abschnitt	5	0	2	<b>7*</b>	1
5. Abschnitt	4	1	<b>6*</b>	3	1

Tabelle 10-1 Bildauswahl Häufigkeiten Mozarttest

Legende: **Fettgedruckt:** Bild, das am häufigsten ausgewählt wurden, \* im eigentlichen Textabschnitt in der WP verwendetes Bild.

Bildauswahl Ø Note	Bild 1	Bild 2	Bild 3	Bild 4
1. Abschnitt	<b>1,53*</b>	<u>2,73</u>	2,53	3
2. Abschnitt	2,2	2,6	<b>1,8*</b>	<u>2,67</u>
3. Abschnitt	2*	2,6	<u>2,13</u>	3,4
4. Abschnitt	1,8*	1,87	2,2	<u>2,2</u>
5. Abschnitt	2,93	2,73	<u>2,47*</u>	3,07

Tabelle 10-2 Durchschnittliche Bewertung der Bilder mit Schulnoten

Legende: **Fettgedruckt**: Bild, das am häufigsten ausgewählt wurden, \* beste Durchschnittsnote, Unterstrichen: im eigentlichen Textabschnitt in der WP verwendetes Bild.

Die Auswertungen zeigen, dass dreimal das im tatsächlichen Wikipedia-Text verwendete Bild ausgewählt wurde, obwohl es nur einmal die beste Durchschnittsnote erhalten hatte. Drei Bilder wurden überhaupt nie ausgewählt (3. Abschnitt Bild 1 und Bild 4, 4. Abschnitt Bild 2, siehe dazu Anhang 2). In drei von fünf Fällen wurde das mit der besten Durchschnittsnote bewertete Bild am häufigsten ausgewählt. Die pro Abschnitt am häufigsten ausgewählten Bilder bekamen zusammen 46:29 Stimmen vs. alle anderen zusammen.

Da aber nicht in allen Fällen das mit der besten Durchschnittsnote bewertete Bild gewählt wurde, werden die einzelnen Begründungen für die Auswahl genauer untersucht.

In 17 von 65 Fällen wurde *nicht* das von der jeweiligen Testperson am besten bewertete Bild ausgewählt (Fall 1, siehe Tabelle 10-3).

Dafür wurden insgesamt 30 Begründungen angegeben.

Begründung	Anzahl	Prozent
Bild passt am besten zum Text	8	26,66%
Bild gefällt mir am besten	0	0%
Mir war es egal, alle Bilder gleich gut/schlecht	2	6,66%
Bildgegenstand ist der wichtigste im Abschnitt	6	20%
Bild ergänzt den Text gut	7	23,33%
Bild repräsentiert den Text gut	3	10%
Bild ist besonders aussagekräftig	2	6,66%
Außergewöhnliches Bild	1	3,33%
Ich mag grundsätzlich diese Art von Bildern	1	3,33%

*Tabelle 10-3 Auswahl der Begründungen Fall 1.*

In den restlichen (38 von 65) Fällen wurde das am besten bewertete Bild oder – im Falle mehrerer gleicher Wertungen – eines der am besten bewerteten Bilder ausgewählt (Fall 2, siehe Tabelle 10-4)

Es wurden insgesamt 119 Begründungen dafür angegeben.

Begründung	Anzahl	Prozent
Bild passt am besten zum Text	40	33,6%
Bild gefällt mir am besten	14	11,7%
Mir war es egal, alle Bilder gleich gut/schlecht	2	1,68%
Bildgegenstand ist der wichtigste im Abschnitt	9	7,56%
Bild ergänzt den Text gut	18	15,12%
Bild repräsentiert den Text gut	18	15,12%
Bild ist besonders aussagekräftig	14	11,76%
Außergewöhnliches Bild	3	2,52%
Ich mag grundsätzlich diese Art von Bildern	1	0,84%

*Tabelle 10-4 Auswahl der Begründungen Fall 2.*

Für alle 65 Fälle zusammen verteilen sich die Begründungen wie folgt:

Begründung	Anzahl	Prozent
Bild passt am besten zum Text	48	32,2%
Bild gefällt mir am besten	14	9,4 %
Mir war es egal, alle Bilder gleich gut/schlecht	4	2,68%
Bildgegenstand ist der wichtigste im Abschnitt	15	10,07%
Bild ergänzt den Text gut	25	16,78%
Bild repräsentiert den Text gut	21	14,09%
Bild ist besonders aussagekräftig	16	10,74%
Außergewöhnliches Bild	4	2,68%
Ich mag grundsätzlich diese Art von Bildern	2	1,34%

*Tabelle 10-5 Auswahl der Begründungen bei allen Bildern*

Insgesamt wurden 149 Begründungen ausgewählt. 48 Mal wurde die Begründung „Bild passt am besten zum Text“ (32,2%) verwendet. Auf Platz zwei und nur halb so oft wurde ausgewählt „Bild ergänzt den Text gut“ (16,78%). Damit ist die ergänzende Funktion in dieser Studie ein wichtiger Aspekt. In den 17 Fällen, in denen nicht das mit der besten Note bewertete Bild ausgewählt worden ist (Fall 1) wurde ebenfalls am häufigsten die Begründung „Bild passt am besten zum Text“ verwendet (26,66%). Auch die zweithäufigste Begründung war hier wie im Gesamtergebnis „Bild ergänzt den Text gut“, allerdings mit viel geringerem Abstand zu Platz eins (23,33 %). Die Begründung „Bild gefällt mir am besten“ wurde im Fall 1 konsequenter Weise nie verwendet.

Begründung	Prozent [Fall 1]	Prozent [gesamt]	Prozent [Fall 2]
1	26,66% [1]	32,2% [1]	33,6%
2	0% [7]	9,4 % [6]	11,7%
3	6,66% [5]	2,68% [7]	1,68%
4	20% [3]	10,07% [5]	7,56%
6	23,33% [2]	16,78% [2]	15,12%
7	10% [4]	14,09% [3]	15,12%
8	6,66% [5]	10,74% [4]	11,76%
9	3,33% [6]	2,68% [7]	2,52%
10	3,33% [6]	1,34% [8]	0,84%

Tabelle 10-6 Ergebnisse aller drei Ereignisse

### 10.8.3 Folgerungen aus der Studie

Bilder werden nach diesen Ergebnissen danach ausgewählt, wie gut sie zum Text „passen“. Darüber hinaus wird der ergänzende Aspekt eines Bildes betont. Diese Kriterien werden den Ergebnissen zu Folge zumindest teilweise über die rein subjektive Bewertung der Bildqualität (hier ausgedrückt durch die eingangs vergebene Schulnote für die einzelnen Bilder) gestellt. Hinsichtlich der Fragestellung dieser Arbeit ist das eine Bestärkung, dass die passende Bildauswahl nicht nur rein über die Suche nach dem „besten“ Bild sondern nach dem am besten „passenden“ Bild erfolgt.



## 10.9 Anforderungen an ein Ordnungssystem für Bilder

Um passende Bilder auswählen zu können ist es – so zeigte auch die genannte Studie, in der dieses Instrumentarium noch nicht bestand und deshalb keine weiteren Aussagen, dazu, was konkret der Nutzer unter „passend“ versteht, gemacht werden konnten – notwendig, die Eigenschaften von Bildern formal zu erfassen. Deshalb wurde ein umfangreiches System zur Klassifikation von Bildern erstellt (gemeinsam mit [Schubart, 2012]). Dieses System findet in mehreren Arbeiten [z.B. [Knäusl und Schubart, 2013; Schubart, 2013] Anwendung. Mit Hilfe der entstandenen Klassifikation lassen sich Bilder nach syntaktische, semantischen, pragmatischen und emotionalen Gesichtspunkten beschreiben (vollständige Klassen siehe Anhang 2).

## 10.10 Fazit

Zusammenfassend ist festzuhalten, dass die Funktionen von Bildern generell unterschiedlich sein können [Levin, 1981]. Aufgrund der auf dem jeweiligen Bildverständnis basierenden Interpretationsweise von Bildern gelten beim Information Retrieval die Aspekte, die der *cognitive viewpoint* [Ingwersen und Järvelin, 2005] beschreibt: Abhängig von Erkenntnisziel, inhaltlichem Vorwissen und allgemeiner Wissensstruktur des Nutzers verändert sich Rezeption und damit auch Funktion(smöglichkeit) eines Bildes. Studien zeigen, dass Bilder abhängig von ihrer Funktion ausgewählt werden [Knäusl et al., 2012-1; Knäusl und Schubart, 2013]. Die hier durchgeführte Untersuchung bestätigt, dass nicht nur die Qualität und die sachliche Richtigkeit für die Auswahl eines Bildes verantwortlich sind, so auch [Schubart, 2013].

Damit wurde gezeigt, dass die Bildauswahl ebenfalls situations- und nutzerrelativ ist. Die Berücksichtigung von Bildern abhängig von der Suchaufgabe wird in einer weiteren Studie untersucht.



## 11 Nutzer-Präferenzen in Abhängigkeit vom Nutzungsszenario

Wie gezeigt wurde, existieren unterscheidbare Nutzungsszenarien (siehe Kapitel 5). Für Teilkomponenten dieser Szenarien wurden in der Literatur bereits Einflüsse auf das Suchverhalten der Nutzer gezeigt (siehe u.a. [Yue et al., 2013; Yan et al., 2011, Calderon, 2012; White et al., 2009; Kelly, 2002; Marchionini, 1989; Wildemuth, 2004; Jochmann, 2010; Moshfeghi und Jose, 2013; Poddar und Ruthven, 2010; Axelrod, 2004; Moshfeghi, 2004; Nahl und Bilal, 2007; Kuhlthau, 2004]).

Es wurde weiterhin besprochen, dass die Berücksichtigung von Nutzerszenarien ein nachvollziehbar probates Mittel ist, um durch entsprechende Unterstützung des Systems die Nutzerzufriedenheit und den Erfolg zu steigern (siehe Kapitel 7.2).

Ob eine dem jeweiligen Nutzungsszenario entsprechend adaptiv gestaltete Informationspräsentation in der Wikipedia sinnvoll ist, hängt weiter davon ab, ob tatsächlich abhängig von den Szenarien eine unterschiedliche Informationsrezeption der Nutzer nachzuweisen ist.

Da diese Frage in der Literatur offen bleibt, wird sie hier in zwei empirischen Studien untersucht.

Der Fokus liegt dabei darauf, konkrete Zusammenhänge zwischen den Faktoren, die das Nutzungsszenario bilden, und dem jeweiligen Suchverhalten der Nutzer zu untersuchen. Nachfolgende Grafik (Abbildung 11-1) veranschaulicht die Überlegung zu den unterschiedlichen Rezeptionsprozessen. Hier werden zwei Suchverläufe dargestellt, wie sie beim IR in der Wikipedia vermutet werden (zeitlicher Verlauf wird durch die Pfeilrichtung symbolisiert). Während Nutzer 1 (repräsentiert durch den Verlauf der roten Pfeile) zunächst vollständig das Inhaltsverzeichnis liest, dann weiter die Überschriften berücksichtigt, bis er offensichtlich einen relevanten Text gefunden hat, nach dessen Lektüre er noch ein Bild betrachtet und dann die Seite verlässt, geht der zweite Nutzer (blaue Pfeile) selektiver vor, studiert nur einen Teil des Inhaltsverzeichnis, springt dann zu einer Infografik und verlässt dann ebenfalls die Seite. In den nachfolgenden empirischen Studien steht genau dieser situative Auswahlprozess im Fokus des Interesses.



Abbildung 11-1 Ausschnitt aus Wikipedia: Regensburg, abgerufen 13. Dezember 2013

Den wichtigsten Stellenwert nehmen dabei als ursächliche Faktoren (also Gründe für die oben benannten unterschiedlichen Vorgehensweisen) die verschiedenen Aufgabentypen (hier v.a. hinsichtlich der zwei Unterscheidungen „inhaltliches Ziel“ und „Motivation“, siehe ab Kapitel 5) ein. Dabei wird auch wiederum Bezug genommen zu der Fragestellung von [Marchionini, 1989], diesmal speziell im Falle der Suche in der deutschsprachigen Wikipedia:

„3. What are the relationships among user, task and search pattern?“ [Marchionini, 1989, S.3]

Cutrell und Guan [Cutrell und Guan, 2007] leisteten einen allgemeinen Beitrag zu dieser Fragestellung: Sie zeigten, dass die vom Nutzer bevorzugte Textlänge zwischen den verschiedenen Tasktypen variiert. Liegt also ein präzise formuliertes/formulierbares Informationsbedürfnis vor, dann präferierte der Nutzer kürzere Textpassagen, wogegen ein Nutzer, der eher explorativ navigierte, auch längere Textstellen las. Gegenstand der Untersuchungen von [Cutrell und Guan, 2007] waren allerdings die Ergebnisseiten von Suchmaschinen. Im Folgenden wird konkret die Nutzung der Inhaltselemente in der deutschsprachigen Wikipedia untersucht.

Als ein zweiter wesentlicher Aspekt werden die emotionalen Reaktionen der Nutzer und ihre generelle emotionale Lage berücksichtigt, da Emotionen hier sowohl Zielgröße als auch ursächlich sind (siehe Kapitel 5.6.7).

Ziel der folgenden Studien ist es, die Präferenzen für bestimmte *Inhaltselemente* und die Informationsrezeption bei der Suche in der deutschsprachigen Wikipedia abhängig vom Nutzungsszenario durch implizites und explizites Feedback zu den präsentierten Suchergebnissen zu untersuchen. Die dabei verwendete Kombination von explizitem und implizitem Feedback verbessert die Genauigkeit der Interpretation des Nutzerverhaltens. Darüber hinaus ist von Interesse, ob es einen Zusammenhang zwischen der Zufriedenheit der Nutzer und dem Maße, in wie weit sie die Informationspräsentation als *angemessen* empfunden haben, gibt.

Da in einer Feldstudie die für die Untersuchungen dieser Arbeit benötigten Daten (siehe dazu auch Kapitel 11.4, Auswahl der Methoden) nicht erhoben werden können, wird eine Laborstudie durchgeführt. Bei einer Laborstudie können durch die im Vergleich zu einer Feldstudie erhöhten Möglichkeiten zur Kontrolle Störfaktoren besser ausgeschlossen werden. Allerdings ist bei einer Laborstudie immer zu berücksichtigen, dass zwar – durch bessere Kontrolle der Störfaktoren – die interne Validität der Daten tendenziell höher ist als im Feld, allerdings muss dies dahingehend eingeschränkt werden, dass eventuelle Rückschlüsse auf ein Verhalten der Personen im Feld weniger valide sind (externe Validität) [Schnell, 2008]. Eine Laborstudie wird hier außer aus Gründen der Methodenwahl auch deshalb bevorzugt, da es sich bei den Nutzungsszenarien um ein mehrdimensionales Konstrukt handelt, das ohnehin schon relativ schwierig zu kontrollieren ist (wer hat welches Vorwissen, Vorlieben, Auffassungsvermögen etc.), so dass ein Maximum an Kontrolle notwendig ist, um interne Validität herzustellen.

Eine Berücksichtigung der *externen* Validität erfolgte vor allem durch die Art und Weise der Aufgabenstellung, siehe dazu Kapitel 11.2.3.

Die Testpersonen hatten die Aufgaben unter den vorgegebenen Bedingungen (siehe dazu 11.2.1 bis 11.2.3) zu erfüllen. Dabei wurden ihre Aktionen mit Hilfe verschiedener

Methoden protokolliert. Aus den so erhaltenen Daten wurden Erkenntnisse gewonnen, ob und wenn ja welche spezifischen Strukturelemente (Bilder/ Infoboxen/ Überschriften/etc.) berücksichtigt verwendet werden, um zum vorgegeben Such-Ziel zu gelangen. Dieses *Implicit Feedback* bedeutet, dass aus der Interaktion des Nutzers mit dem System und den Inhalten Rückschlüsse auf deren situative Relevanz gezogen werden [White und Kelly, 2006; Kelly und Fu, 2007]. Zusätzlich wurden die Testpersonen um explizites Feedback gebeten.

## 11.1 Aufbau

In den Studien werden informationsorientierte Suchtasks nach [Broder, 2002] untersucht, die mit Hilfe des Modells weiter nach den in Kapitel 6.2.1 beschriebenen Variablen unterteilt werden.

Methodenauswahl und Aufbau der Studie werden im Folgenden beschrieben.

### 11.1.1 Methodenauswahl – Emotionen

Die emotionale Lage des Nutzers wird als ein wesentlicher Einflussfaktor erfasst. Wesentlich deshalb, da gerade bei der Freizeitsuche die Stimmung und die emotionale Lage eine wichtige Rolle als Zielgröße spielen [Elsweiler und Wilson, 2010] und zweitens die Suche nach Informationen auch eine affektive Komponente hat [Nahl und Bilal, 2007; Kuhlthau, 1991] sowie Wechselwirkung zwischen Suchprozess und Emotionen nachgewiesen werden konnten [Moshfeghi und Jose, 2013]. Die Berücksichtigung von Emotionen z.B. mittels implizitem Feedback [Feild et al., 2010] kann dabei helfen, auf Frustration des Nutzers zu reagieren.

### 11.1.2 Definition und Begrifflichkeit von Emotion – Modelle

Für die Definition, Erfassung und Auswertung von Emotionen stehen verschiedene Möglichkeiten zur Verfügung. Im Folgenden werden diese Möglichkeiten und die wichtigsten Theorien, Modelle und Methoden zum Thema Emotionen besprochen.

*"Jeder weiß, was eine Emotion ist, bis er gebeten wird, eine Definition dafür zu geben."*

James Russell und Ernst Fehr, zitiert nach [Pontes, 2011]

Emotionen zu definieren und vergleichbar zu machen ist schwierig. Will man aber Zusammenhänge herstellen und generelle Aussagen ermöglichen, ist es unverzichtbar, eine einheitliche Definition und Interpretation bereit zu stellen.

[Kleinginna und Kleinginna, 1981] haben über 90 Definitionen für Emotionen in zehn Kategorien zusammengestellt (siehe Tabelle 11-1). Unterscheidungen sind dabei nach vielen Kriterien möglich, wie [Kleinginna und Kleinginna 1981] zeigen: Handelt es sich dabei um neurophysiologische Vorgänge, ein Zustand mit bestimmten Gefühlen oder sind Emotionen rein die Benennung und Bewertung einer Situation? Je nach Interpretationsansatz verändert sich die Plausibilität der verschiedenen Sichtweisen.

*Tabelle 11-1 Zusammenstellung von Emotionen nach [Kleinginna und Kleinginna, 1981, S.17]*

Kategorie	Beispiel für eine Definition
Affektiv (Gefühl)	... ein Zustand, in dem das Individuum Gefühle erlebt
Physiologisch	... Verhalten, primär durch viszerale Reaktionen beeinflusst
Kognitiv	... schließen Bewertungen ein
Ausdruck	... Emotionen und ihr Ausdruck bilden eine existentielle Einheit
Multiple Aspekte	... komplexes Konzept mit neurophysiologischen, muskulären und phänomenologischen Aspekten
Motivational	... motivationale Zustände
Externe Reize	... werden allgemein durch externe Reize ausgelöst
Abgrenzend	... primär Gefühlszustand, während Motivation ...
Adaptiv	... Signal, das den Organismus ... vorbereitet
Disruptiv	... führt zur Unterbrechung der üblichen Verhaltensmuster

Wie auch bei anderen terminologischen Grundsatzdiskussionen gelten hier zwei Grundsätze:

- 1) Am sinnvollsten legt man eine Definition fest, die erst einmal nur für den Forschungsbereich, in dem sie verwendet wird, gültig ist
- 2) Alle Definitionen und Herleitungen sollten überprüft werden und dann aus pragmatischer und wissenschaftlich logischer Sichtweise die Aspekte verwendet werden, die sinnvoll, hilfreich und schlüssig sind (für den aktuellen Forschungsbedarf).

In dieser Arbeit geht es darum, Auskünfte über die aktuellen Emotionen der Nutzer zu erhalten, um Einflüsse auf ihr Suchverhalten zu erkennen und zweitens ihre emotionale Reaktionen auf die Suche und die Inhalte verfolgen zu können (Veränderung der

Emotion bzw. Erreichen eines bestimmten emotionalen Zustands als Zielgröße und Emotion als Einflussfaktor). Dabei ist vor allem eine Vergleichbarkeit der gemessenen Emotions-Werte notwendig. Es wird also Ausdruck, Darstellung und Definition von Emotionen benötigt, die einheitlich verständlich sind, mit vorhanden Messmethoden mess- und darstellbar. Darüber hinaus sollen die Emotionen so erfasst werden können, dass mathematische Operationen auf die Ergebnisse angewendet werden können, da Zusammenhänge zwischen den anderen Variablen des Modells (Kapitel 6) untersucht werden.

Jeder Mensch *zeigt* mehr oder weniger Emotionen in Gestik, Mimik, verbalen Äußerungen und auch physiologischen Zeichen wie Blutdruck, Gesichtsfarbe, Atmung, Gehirnaktivität und muskulärer Anspannung. Diese Zeichen allgemein und vergleichbar zu fassen, zu messen und sogar in für eine Maschine verwertbare Informationen umzuwandeln ist allerdings schwierig, da die Ausdrucksweisen emotionaler Reaktionen erst einmal keine direkt verwertbare Information enthalten.

Emotionen haben weiterhin einen starken Einfluss auf das Verhalten eines Menschen. Durch einen Auslöser erzeugt können sie das Verhalten oder das Interesse eines Menschen schnell verändern und steuern [Scherer, 2005]. Somit könnte man auch das Verhalten der Menschen verwenden, um auf Emotionen rück zu schließen. Dies entspricht nach [Kleinginna und Kleinginna, 1981] etwa einer physiologischen Definition. Für die vorliegende Arbeit ist die *kognitive* Sichtweise plausibel. Denn diese schließt eine *Bewertung* mit ein [Kleinginna und Kleinginna, 1981]. Da in dieser Arbeit bewusst von objektiven Konzepten, vor allem in Hinsicht auf Relevanz und Bewertung, Abstand genommen wird und der Fokus auf der kognitiven Sichtweise liegt, gilt das auch für das Thema der Emotionen.

Nach der *kognitiven Emotionstheorie* [Lopatovska, 2009] werden Emotionen durch kognitive Aktivität ausgelöst. Das bedeutet, irgendetwas, mit dem sich der Mensch in diesem Moment befasst, führt dazu, dass eine Emotion entsteht. Der Gegenstand der kognitiven Aktivität kann absichtlich oder unabsichtlich im Fokus der Person stehen, das spielt dabei keine Rolle. Nach [Schmidt-Atzert, 1996] ist darüber hinaus aber auch die *subjektive Bewertung* des Individuums ausschlaggebend dafür, welche Emotionen durch



etwas ausgelöst werden. Dies ist vor allem plausibel, berücksichtigt man die Erinnerungen und Assoziationen mit Dingen, Ereignissen, Gerüchen, Musik und dergleichen, die sehr subjektiv und individuell sind. Daraus lässt sich die Vermutung ableiten, dass die Interaktion mit den Inhalten der Wikipedia bei den Nutzern unterschiedliche Emotionen bzw. die affektiven Reaktionen auslösen kann.

Der *nicht-kognitive* Ansatz wiederum besagt, dass Emotionen unabhängig von der Kognition entstehen und vielmehr dazu da sind, um körperliche Funktionalitäten zu steuern und zu optimieren – eher mit dem Instinkt (weglaufen bei Angst) zu vergleichen. Danach ist das emotionale Erleben ausschlaggebend dafür, wie bestimmte Dinge bewertet werden [Schmidt-Atzert, 1996] und nicht umgekehrt.

Grundsätzlich ist aber anzunehmen, dass kognitive Aktivitäten verschiedener Art bei der Entstehung von Emotionen eine wesentliche Rolle spielen [Lapotovska, 2009]. Das kann sein, dass man etwas als positiv Abgespeichertes wiedererkennt, dass Schlüsselreize angesprochen werden wie etwa beim „Kindchenschema“, dass etwas einfach den Geschmack trifft oder – im negativen, Angst (durch erlerntes *oder* instinktives Verhalten), negative Erinnerungen, Ärger und mehr.

Es bleibt weiterhin bezüglich des Anwendungskontextes dieser Arbeit festzuhalten, dass emotionale Reaktionen auf Stimuli wie Bilder oder auch andere Inhalte beim Information Retrieval sowohl intersubjektiv und nachvollziehbar als auch subjektiv und damit nicht objektiv nachvollziehbar sein können. Entsprechend der *appraisal theory* [Scherer und Shorr, 2001] wird hier davon ausgegangen, dass nicht nur die Emotionen die Kognition beeinflussen sondern umgekehrt genauso die Kognition einen Einfluss auf die jeweiligen Emotionen des Individuums hat.

### 11.1.2.1 Benennung und Beschreibung von Emotionen

Eine weitere Herausforderung ist die Benennung und Beschreibung von Emotionen. Neben der Festlegung einer Sichtweise auf Emotionen (hier: kognitive Sichtweise) müssen Emotionen benannt werden, um explizit gemacht werden zu können.

*„Emotion schemes, the most frequently occurring emotion experiences, are dynamic emotion cognition interactions that may consist of momentary/ situational responding or enduring traits of personality that emerge over developmental time. Emotions play a critical role in the evolution of consciousness and the operations of all mental processes.“* [Izard, 2009, S.4]

Grundsätzlich besteht unter Psychologen Einigkeit darüber, dass Menschen immer auf ihr direktes Umfeld emotional reagieren und auch darüber, dass die ersten und ursprünglichsten Reaktionen affektiver Art sind [Machleit und Eruglo, 2000]. Emotionen beeinflussen somit das Handeln und Denken von Menschen zu jeder Zeit.

Es ist allerdings nicht trivial, diese Emotionen „beim Namen zu nennen“. In der Regel nimmt man seine eigenen Emotionen zwar immer wahr, sie aber konkret zu beschreiben gestaltet sich schwieriger. Die Benennung von Emotionen entsteht in der *„Ausdifferenzierung kontextualisierter Interaktionsverläufe“* [Beumer et al., 2011, S. 92]. Sie wird durch das Umfeld bzw. andere Personen vermittelt, damit sinkt die Treffsicherheit und eine Distanzierung bzw. Abstrahierung von dem konkret Erlebten findet statt [Beumer et al., 2011]. Diese Benennung erlernt zwar jeder in einem sozialen und von Bezugspersonen geprägten Umfeld aufwachsende Mensch, allerdings handelt es sich dabei nicht um konstantes Wissen sondern dieses Erlernen wird durch Erfahrung beständig erweitert [Beumer et al., 2011]. Damit ist emotionales Erleben (und die Trennung von eigener Interpretation und erlernter Intention) immer einer Entwicklung unterworfen und somit die Benennung und der Austausch über Emotionen nicht zwingend intersubjektiv einheitlich und nachvollziehbar.

Schwierig ist es, selbst wenn man seine eigenen Emotionen beschreiben kann, eine einheitliche Vorstellung und Interpretation der verwendeten Begrifflichkeiten und der verschiedenen Dimensionen wie etwa der Intensität zu erzeugen.

Im Wesentlichen gibt es bei der Benennung und Beschreibung von Emotionen zwei Ansätze. Der diskrete Ansatz (siehe dazu [Lopatovska, 2009]) geht davon aus, dass es ein „Inventar“ an Basisemotionen gibt, die allen zu Emotionen fähigen Individuen möglich sind und die auch immer von anderen eindeutig erkannt werden (z.B. Angst, Wut oder Freude). Alle weiteren Emotionen setzen sich anteilig aus diesen Basisemotionen zusammen.

Der kontinuierliche Ansatz dagegen geht von mehreren Dimensionen aus, die Emotionen ausmachen. In den wesentlichsten Forschungen zu diesem Thema wird in der Regel von zwei oder drei Dimensionen ausgegangen (siehe dazu [Wundt, 1896; Bradley und Lang, 1994; Osgood, 1952; Russell und Mehrabian, 1977]). Russel und Mehrabian zeigen, dass drei bipolare Dimensionen ausreichen, um das Spektrum der menschlichen Emotionen beschreiben zu können, basierend auf den Untersuchungen von [Wundt, 1896], der die These aufstellte, dass emotionale Reaktionen durch bestimmte Stimuli gut anhand von drei grundsätzlichen Dimensionen beschrieben werden können: Lust, Spannung und Beruhigung. Auch andere Autoren (z.B. [Osgood, 1952; Osgood et al., 1957; Russel und Mehrabian, 1977 bzw. Russell, 1979]) bestätigen die drei – unabhängigen – Dimensionen zur Erfassung emotionaler Zustände: *pleasure – displeasure, degree of arousal, and dominance-submissiveness*.

Grundsätzlich zeigen Untersuchungen, dass die Verbalisierung von Emotionen trotz der Übereinkunft in den Theorien aus verschiedenen Gründen wie unterschiedlicher Erfahrung, Abstraktion und Eigenwahrnehmung [Beumler et al., 2011] schwierig ist. Die einheitliche Darstellung und Beschreibung bleibt eine Herausforderung. Es existieren verschiedene Ansätze zur Lösung des Problems, die im Folgenden besprochen werden, um eine geeignete Auswahl für die empirische Anwendung zu treffen.

#### **11.1.2.2 Basisemotionen nach Ekman**

Der nicht-kognitiven Sichtweise entsprechend gibt es verschiedene Autoren, die eine Grundausstattung von Basisemotionen konstatieren (vgl. [Izard, 1987] oder [Ekman, 1992]). Bestärkend für diese Theorie ist, dass alle Kulturen gewisse Emotionen kennen wie z.B. Trauer, Angst und Freude. Die jeweiligen Mimiken sind vergleichbar. Auch bei kleinen Kindern sind die Gesichtsausdrücke eindeutig erkennbar, noch bevor aktiv

erlerntes Verhalten möglich ist. Selbst bei Primaten wurden ähnliche Mimiken festgestellt [Lopatovska, 2009].

Darüber hinaus lösen die sog. Basisemotionen vergleichbare Reaktionen aus (wie z.B. Flucht bei Angst oder erhöhte Alarmbereitschaft durch Ausschüttung entsprechender Hormone). Auch Charles Darwin [Darwin, 1872/1965] ging davon aus, dass Emotionen universell und angeboren sind. Unklar ist allerdings, wie aus der „Mischung“ der Basisemotionen dann das restliche Spektrum affektiver Zustände entstehen kann. Darüber hinaus sind die Basisemotionen in den Theorien unterschiedlich.

Ekman hat sieben Basisemotionen identifiziert [Ekman, 1992]: Freude, Wut, Ekel, Furcht, Verachtung, Traurigkeit und Überraschung. Seines Erachtens sind diese Emotionen universell und kulturell unabhängig erkennbar. Er schließt nicht aus, dass erlerntes und sozial erwünschtes Verhalten die Ausdrucksweise dieser Basisemotionen beeinflusst, allerdings sind diese „Täuschungsmanöver“ in der Regel nicht so stark, dass die eigentlichen Reaktionen nicht mehr erkennbar sind.

Mit Hilfe des *Facial action coding systems* ordnet Ekman die jeweiligen Mimiken den Emotionen zu. So lassen sich zum Beispiel auch gekünstelte Gesichtsausdrücke „entlarven“ [Ekman und van Friesen, 1978]. Bei dieser Methode fällt die Selbstauskunft und damit Benennung von Emotionen weg. Für die Anwendung allerdings bedarf es Übung und Erfahrung, weshalb diese Methode hier nicht zum Einsatz kam.

#### **11.1.2.3 Das OCC-Modell [Ortony, Clore und Collins]**

Das OCC-Modell ist ein wichtiges Modell in der Informationsverarbeitung. Bei der Arbeit mit KI-Systemen wird es häufig verwendet, um Emotionen zu modellieren. Ortony et al. [Ortony et al., 1988] versuchen darin, vor allem zwei wesentlichen Fragen nachzugehen und so die schwierige Darstellbarkeit von Emotionen zu verbessern. Sie betonen vor allem das Ziel, die Emotionen komplexer darzustellen als „nur“ eine Klassifizierung nach positiv und negativ [Ortony et al., 1988].



Wikipedia ist das Modell allerdings nicht stark genug auf diese zwei Aspekte fokussiert.

#### 11.1.2.4 Das Circumplex-Modell von Russel

Russel [Russell, 1980] greift mit seinem Circumplex-Modell den Gedanken der verschiedenen, voneinander unabhängigen Emotionsdimensionen auf. Er verwendet dabei allerdings nur die zwei Dimensionen Valenz und Erregung.

In seinem Modell ordnet er die emotionalen Konzepte kreisförmig an, wobei die Ausprägungen festgelegt sind: pleasure 0°, excitement 45°, arousal 90°, distress 135°, displeasure 180°, depression 225°, sleepiness 270° und relaxion 315°. [Russell, 1980] weist 28 englischsprachige Adjektive, die allgemein zur Beschreibung emotionaler Zustände herangezogen werden, auf dem Kreis den Konzepten zu. Die Mitte bildet sozusagen der „emotionale Nullpunkt“ (siehe Abbildung 11-3).

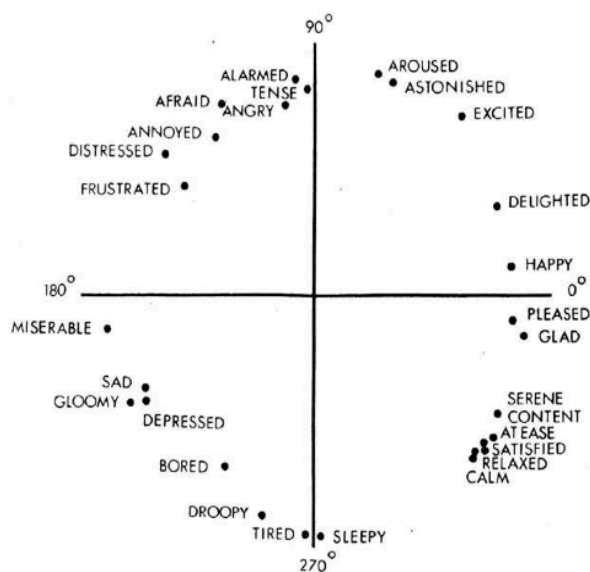


Abbildung 11-3 Circumplexmodell [Russell, 1980, S.7]

Das Circumplexmodell kann für verschiedene Zwecke adaptiert werden, wie zum Beispiel die emotionale Reaktion auf Produkte [Desmet und Hekkert, 2007, siehe Abbildung 11-4]. Durch die Kreisförmige Anordnung können die Benennungen der Emotionen in entsprechender Gewichtung zu den verwendeten Wörtern getätigt werden.

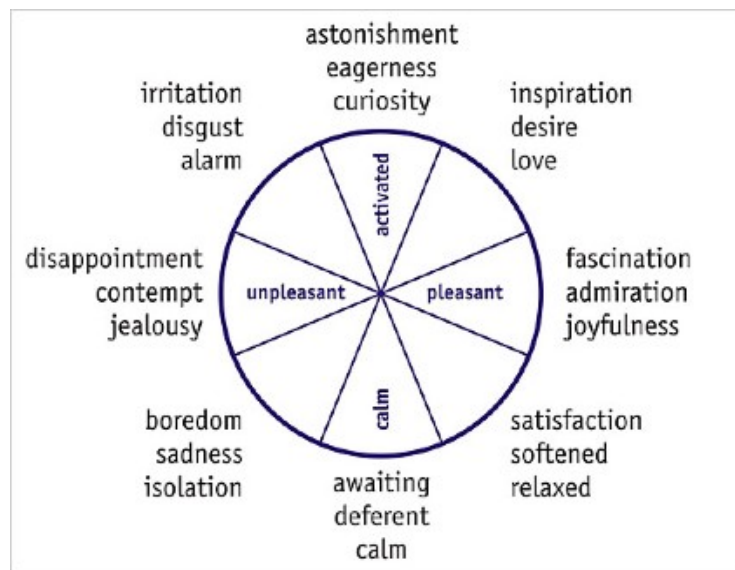


Abbildung 11-4 Circumplexmodell [Desmet und Hekkert, 2007], basierend auf [Russell, 1980]

Durch vorgegebene Adjektive ist eine einheitliche Beschreibung gesichert. Für die Selbstausskunft der Probanden zu ihrer emotionalen Lage ist dies eine gute Grundlage, da die Personen keine eigenen, eventuell unterschiedlich konnotierten Wörter verwenden müssen. Lediglich die einheitliche Interpretation oder Verständnis der Adjektive ist ein Schwachpunkt dieser Methode. Vorteil ist, dass – anders als beim semantischen Differential (siehe 11.1.3.1) – nicht viele Skalen verwendet werden müssen, sondern durch die kreisförmige Anordnung diese Dimensionen gleichzeitig erfasst werden können.

### 11.1.3 Subjektive Methoden um Emotionen zu erfassen

Zur Beschreibung und Benennung von Emotionen gehört nicht nur die terminologische Einigung, sondern Emotionen haben noch weitere Eigenschaften wie ihre Intensität, den Auslöser, die physischen Reaktionen und den Verlauf. Einige dieser Eigenschaften sind durch Beobachtung erfassbar (bzw. durch bestimmte Messmethoden), andere müsste die Person selbst beschreiben. Durch unterschiedliche subjektive und objektive Erfassungsmethoden können diese Aspekte beobachtet werden.

Das eigentliche Gefühl einer Person ist zunächst nur ihr selbst zugänglich (anders als die jeweiligen, sichtbaren Reaktionen darauf). Um das eigene emotionale Erleben einer anderen Person mitteilen zu können muss das Gefühl erst „übersetzt“ werden: In

Sprache, Bilder, Fragebögen. Dabei kommt es naturgemäß zu einer gewissen Verzerrung des Erlebten.

Es gibt Methoden zur Erfassung von Emotionen, die entwickelt wurden, um diese Verzerrung so gering wie möglich zu halten. Manche Methoden benutzen dazu Sprache, z.B. [Russell, 1979; Nahl und Bilal, 2007] und manche Bilder, z.B. [Bradley und Lang, 1994].

Eine Methode, um die Intensität von Emotionen (in mehreren Dimensionen) zu erfassen, sind Ratings. Hierbei kommen verschiedene Skalen zum Einsatz, anhand derer versucht werden kann, die unterschiedlichen in 11.1.2 und 11.1.2.1 beschriebenen Dimensionen von Emotionen zu erfassen.

#### **11.1.3.1 Das Semantische Differential**

Das Semantische Differential (*Differential Emotions Scale*) ist ein standardisiertes Messverfahren aus der Psychologie, um die affektive Bedeutung von Objekten oder Begriffen für Personen erfassen zu können [Osgood et al., 1957]. Dazu werden die jeweiligen Begriffe oder Objekte (oder auch Bilder, Musik...) anhand bipolarer Ratingskalen bewertet. Es werden keine konkreten Fragen zu den Objekten verwendet (wie etwa: „Wie hat Ihnen dieses Bild gefallen?“) sondern die Emotionen werden quantifiziert erfasst. Die Ratingskalen beschreiben affektive Faktoren ebenfalls in mehreren semantischen Dimensionen [Osgood et al., 1957; Russell und Mehrabian, 1977; Russell, 1979].

Die Dimensionen *evaluation*, *potency* und *activity*:

- 1) *Evaluation* (Valenzdimension, *pleasure* bei [Russell und Mehrabian, 1977] und [Bradley und Lang, 1994]). In dieser Dimension soll die hedonistische Qualität eines Gegenstands oder Begriffs erfasst werden (angenehm – unangenehm)
- 2) *Potency* (Potenzdimension, *dominance* bei [Russell und Mehrabian, 1977] und [Bradley und Lang, 1994]). Mit Hilfe dieser Skala soll erfasst werden, wieviel Macht und Stärke eine Testperson mit einem Begriff verbindet (stark – schwach) bzw. auch, ob sich der Begriff groß oder klein anfühlt.



- 3) *Activity* (Aktivierungsdimension, *arousal* bei [Russell und Mehrabian, 1977] und [Bradley und Lang, 1994]). In diesem dritten Bereich dient die bipolare Skala dazu, die Erregung der Testperson hinsichtlich eines Gegenstands oder Begriffs zu erfassen: Konnotiert sie den Untersuchungsgegenstand eher als laut und dynamisch oder tendenziell ruhig und leise?

Mit Hilfe dieser drei Dimensionen lassen sich zumindest die „Grundemotionen“ der Menschen darstellen.

In der Praxis gibt es nicht *das eine* semantische Differential. Je nach Verwendungszweck werden die Wortpaare und die Anzahl der Wortpaare variiert. Das zentrale Vorgehen bleibt aber dabei gleich, es werden hauptsächlich fünf- oder siebenstufige Likertskalen auf der bipolaren Basis von Russel [Russell und Mehrabian, 1977] verwendet:

angenehm    1       2       3       4       5       unangenehm

Oder:

ruhig   1       2       3       4       5       6       7       erregt

Alternativ können auch analoge Skalen verwendet werden:

*hässlich* \_\_\_\_\_ *schön*

Diese Skalen können gut in Fragebögen *vor* oder *nach* einem Experiment zum Einsatz kommen. Es müssen nicht immer alle 18 von Mehrabian und Russel [Russell und Mehrabian, 1977] vorgeschlagenen bipolaren Skalen verwendet werden. Grundsätzlich ist das semantische Differential eine einfach anzuwendende Methode, die ein umfassendes Spektrum an affektiven Konnotationen abdeckt. Bei den bipolaren Wortpaaren ist vor allem das Ziel, die Emotionen hinsichtlich eines Stimulus zu beschreiben. Um genaue Auskünfte zu erhalten, sind mehrere Wortpaare notwendig.

Christine Nahl [Nahl, 2007] wendete eine vergleichbare Methode in Telefoninterviews speziell auf die Situation des Information Retrievals an. Sie bat die Testpersonen, Situa-

tionen zu beschreiben, in denen sie eine Informationssuche starten („*troublesome academic life situation*“, „*situation involving scholarship or research*“ etc.). Um die emotionale Komponente dieser Ausgangssituation zu erfassen, sollten die Testpersonen bei jeder Situation angeben, inwiefern sie diese als „*challenging*“, „*important*“, „*confusing*“, „*unfamiliar*“, „*emotional*“ und „*contradictory*“ einschätzten. Dazu wurden jeweils vorgegebene Skalen benutzt, um die Intensität anzugeben. Anhand der Kombination der Ratings der einzelnen Adjektive konnte so ein Gesamteindruck der affektiven Wahrnehmung der Situation der Suche dargestellt werden. Die Mehrdimensionalität der Beschreibung durch die Verwendung der verschiedenen Wörter berücksichtigt, dass emotionales Erleben nicht nur mit einer Dimension (z.B. gut – schlecht) beschrieben werden kann [Wundt, 1896; Osgood, 1952; Russell und Mehrabian, 1977].

In den Untersuchungen der vorliegenden Arbeit wurde allerdings gegen die Methode des Ratings entschieden, da die Nutzer pro Testeinheit sechs Mal nach ihrem emotionalen Zustand gefragt werden sollten und eine sechsfache Angabe der Ausprägung der spezifischen Adjektive ein zu hoher Aufwand (für die Testpersonen) schien.

#### **11.1.3.2 SAM**

Für die empirischen Untersuchungen wurde deshalb eine non-verbale Methode zur Erfassung der Emotionen der Nutzer verwendet, das *Self Assessment Manikin* von Bradley und Lang [Bradley und Lang, 1994], eine stilisierte menschliche Gestalt.

Um die Schwächen der verbalen Erfassung von Emotionen zu vermeiden, wie etwa den relativ hohen Aufwand, viele oder alle der von Russell et al. [Russell und Mehrabian, 1977] vorgeschlagenen 18 Ratings der Emotionen zu erfassen, oder aber auch die Sprachbarriere bei Testpersonen, die unter Umständen entweder nicht der verwendeten Sprache mächtig sind (es gibt die bipolaren Ratings zwar in mehreren validen Übersetzungen, aber das deckt nur einen sehr kleinen Teil aller Sprachen ab) oder nicht lesen und schreiben können (Analphabeten, Kinder...), verwenden Bradley und Lang [Bradley und Lang, 1994] das „*Self-Assessment-Manikin (SAM)*“, um mit Bildern anstelle von Worten die Gefühlslage von Menschen zu beschreiben.

Es kommen hier ebenfalls drei Dimensionen zum Einsatz: *pleasure*, *arousal* und *dominance*. Die drei Dimensionen basieren auch hier auf den Untersuchungen von Wundt [Wundt, 1896].

Die Dimension *pleasure* wird durch den Gesichtsausdruck des Männchens dargestellt, die Dimension *arousal* durch den Magen und die dritte Dimension *dominance* durch die Größe. Jede Dimension ist fünfmal abgebildet, wobei jeweils das entsprechende Detail von Bild zu Bild verändert, um die Stärke der Ausprägung darzustellen (vgl. Abbildung 11-5)

In einer Studie fanden die Autoren heraus, dass Untersuchungen mit SAM vergleichbar genaue Ergebnisse lieferten wie die Untersuchungen mit der Kontrollgruppe, bei der das Semantische Differential mit den 18 bipolaren Skalen verwendet wurde [Bradley und Lang, 1994].

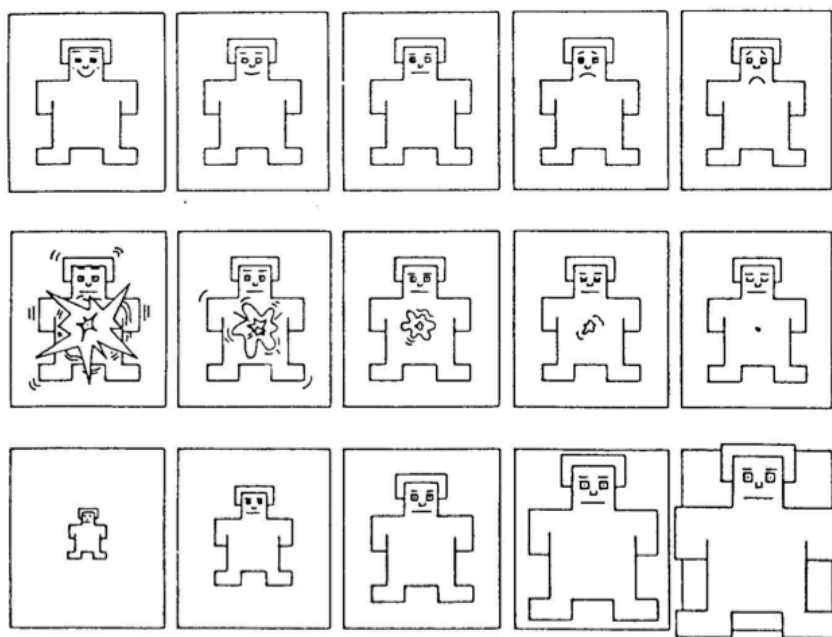


Abbildung 11-5 The Self Assessment Manikin (SAM), [Bradley und Lang, 1994, S. 3]

Das SAM kam in dieser Arbeit zum Einsatz, da so die Selbstausskunft für die Probanden einfacher und schneller möglich war als mit bipolaren Ratings. Durch den Verzicht auf Sprache kann auch die unterschiedliche Interpretation der Adjektive vermieden werden.

#### 11.1.4 Objektive Arten, Emotionen zu erfassen

Um der möglichen Verzerrung, die bei der subjektiven Erfassung von Emotionen entstehen kann, vorzubeugen, gibt es alternativ auch Methoden, die einen objektiveren Ansatz verfolgen. Für die Studien wurde zusätzlich zur Selbstauskunft der Probanden eine objektive Methode zur Erfassung von Emotionen gesucht, um diese (subjektive und objektive) ergänzend anwenden zu können.

##### 11.1.4.1 Klassifikation von Gesichtsausdrücken

Eine verbreitete Methode, mit Hilfe eines Computers zu versuchen, Emotionen zu erkennen, basiert auf der Klassifikation von Gesichtsausdrücken [Picard, 1998; Ekman, 1997]. Mimik ist eine effektive Art und Weise, wie Menschen in der Kommunikation untereinander Informationen über ihren emotionalen Zustand austauschen können. Im Bereich des *Affektive Computing* stellt die automatische Erkennung und Zuordnung von Gesichtsausdrücken eine große Herausforderung dar und wird stark beforscht, sowohl auf dem Gebiet der Bildverarbeitung als auch im Rahmen des *Machine Learning* bis hin zu Untersuchungen der allgemeinen menschlichen Wahrnehmung [Rahulamathavan et al., 2012].

Im wesentlichen funktioniert das Verfahren so, dass zunächst eine Gesichtserkennung (basierend auf Form und Farbe) im Kamerabild durchgeführt wird. Nach bestimmten Mustern werden die einzelnen Gesichtsteile ermittelt und die Partien ausgemessen. Diese Ergebnisse werden klassifiziert und bestimmten Emotionen zugeordnet. Das *facs*<sup>44</sup> von Ekman und Friesen [Ekman und Friesen, 1978] unterscheidet 44 kleinste Bewegungseinheiten im Gesicht, anhand derer Rückschlüsse auf emotionale Regungen gezogen werden können. Dazu zählen z.B. das Hochheben einer Augenbraue, Veränderungen der Mundwinkel oder Rümpfen der Nase.

---

<sup>44</sup> Facial Action Coding System.

**Action Units des Obergesichtes:**

- 1 [Hochheben der inneren Augenbraue \(inner brow raiser\)](#)
- 2 [Hochheben der äußeren Augenbraue \(outer brow raiser\)](#)
- 4 [Zusammenziehen der Augenbrauen \(brow lowerer\)](#)
- 5 [Heben der Oberlider \(upper lid raiser\)](#)
- 6 [Heben der Wangen \(cheek raiser and lid compressor\)](#)
- 7 [Anspannen der Augenlider \(lid tightener\)](#)
- 43 [Schließen der Augen \(Eye closure\)](#)
- 45 [Blinzeln \(blink\)](#)
- 46 [Zwinkern \(wink\)](#)

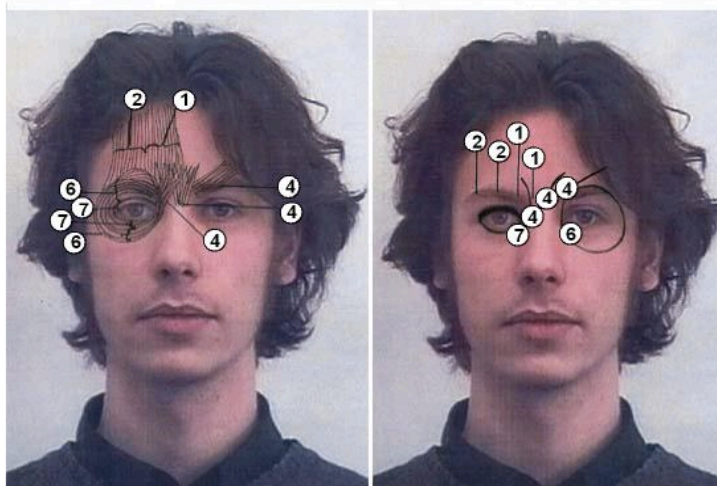


Abbildung 11-6 Klassifikation der Gesichtsbewegungen nach Ekmans Facial Action Coding System [Ekman und Rosenberg, 1997], Quelle: [<http://www.gnosisfacialis.de/Seiten/facs/facs1.htm>] abgerufen am 17. September 2013

Die Gesichtserkennung kann kombiniert werden mit anderen objektiven Verfahren wie Blutdruckmessung oder einer unabhängigen Bewertung von Kontext und Situation (z.B. Darstellung von lustigen Inhalten oder dem Wissen über eine Person, dass sie z.B. gerade eine traurige Nachricht erfahren hat und dementsprechend mit hoher Wahrscheinlichkeit auch ein einer eher negativen emotionalen Verfassung ist [Picard, 1998]).

Vyzas [Vyzas, 1999] zeigt mit Hilfe verschiedener Klassifikationsmethoden, dass einige physiologische Indikatoren zur Analyse des emotionalen und kognitiven Zustands einer Person herangezogen werden können. Das Problem ist allerdings auch hier, dass die Werte schwierig zu verallgemeinern sind und viele Daten aus Laborexperimenten stammen, in denen sich die Testpersonen unter Umständen nicht so verhalten, wie sie es in Wirklichkeit in einer entsprechenden Situation tun würden [Picard, 1998].

#### 11.1.4.2 Elektromyograph

Eine weitere Methode der Emotionserkennung auf Basis der Gesichtsmimik beruht auf der Verwendung eines Elektromyographen (EMG). Die Methode kommt aus der Biomechanik. In diesem Fall spielen Methoden der Bilderkennung keine Rolle, sondern es werden die Spannungswechsel der Gesichtsmuskeln (im Sinne elektrischer Spannung) mit Hilfe von auf der Haut aufgebrachter Elektroden aufgezeichnet.

Auch bei diesem Verfahren [Brown, 1990] ließen sich deutliche Zusammenhänge zwischen bestimmten (Verhaltens-)Mustern der Gesichtsmuskulatur und einzelnen Emotionen erkennen. Diese Ergebnisse wurden betont durch verschiedene Experimente [Brown, 1980], in denen die Selbstausskunft der Versuchspersonen über eine schlechte emotionale Verfassung mit den Aktivitäten der Gesichtsmuskeln (hier: *Korrugator*) verglichen wurden. In diesem Fall wurden die Reaktionen zweier Gesichtsmuskeln protokolliert: *Korrugator* und *Zygomaticus Major*, die als Indikatoren für emotionale Reaktionen gelten [Brown, 1980].

### Information about facial muscles

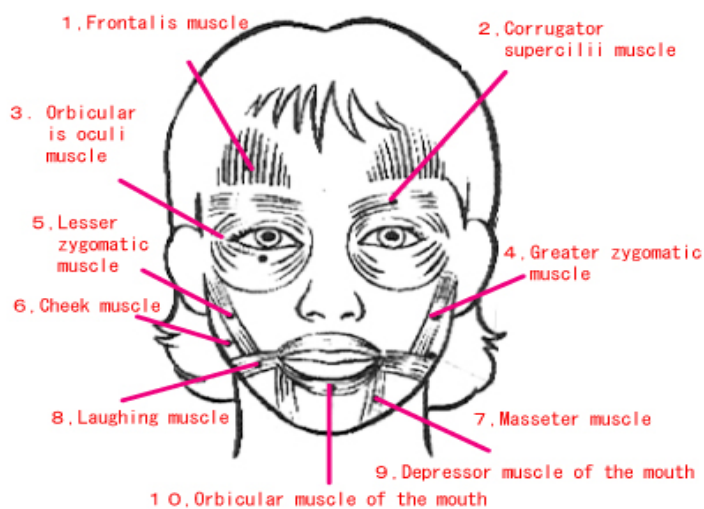


Abbildung 11-7 Overview of facial muscles [source: <http://www.musashi-rie.com/e-kaokin.htm>], abgerufen am 15.Juni 2012

Einen von dieser Methode überzeugenden Versuch zeigt Picard [Picard, 1998] in einem Experiment mit dem Computer Spiel „DOOM“. Für eine kurze Zeitspanne funktionier-

te absichtlich die Software nicht – man sieht deutliche Anspannung im EMG, die mit Pulsschlag und Leitfähigkeit der Haut korreliert (siehe dazu Abbildung 11-8). In den Untersuchungen von [Picard, 1998] zeigte sich die Eignung des EMGs für Untersuchungen zur akuten emotionalen Lage der Versuchspersonen.

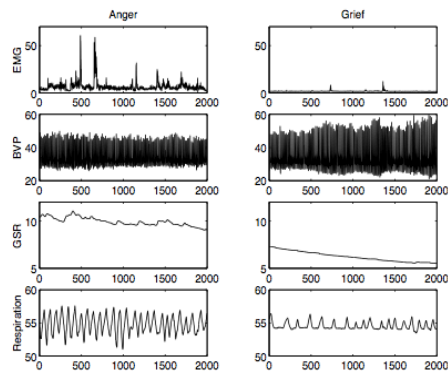


Figure 1: Examples of physiological signals measured from an actress while she consciously expressed anger (left) and grief (right). From top to bottom: electromyogram (microvolts), blood volume pulse (percent reflectance), galvanic skin conductivity (microSiemens), and respiration (percent maximum expansion). All of these signals can be gathered from sensors on the surface of the skin, without any pain or discomfort to the person. These signals were sampled at 20 Hz, using a ProComp system from Thought Technology Ltd. Each box shows 100 seconds of response.

Abbildung 11-8 [Picard, 1998,S.4] Signale des EMG im Vergleich mit Puls und Leitfähigkeit der Haut

Das EMG hat der Bilderkennung gegenüber den Vorteil, dass es schon auf kleinste Bewegungen reagiert, die unter Umständen von einem Beobachter noch nicht wahrgenommen werden. Allerdings erfasst das EMG nur das *Maß* des motivationalen Engagement einer Muskelgruppe; Angst und Wut zeigen zum Beispiel vergleichbare Aktivitäten und sehen bei Verwendung des EMG relativ ähnlich aus. Darüber hinaus ist zu berücksichtigen, dass weitere Störfaktoren möglich sind, wie etwa zusammengekniffene Augen beim Lesen.

Zusammenfassend ist festzustellen, dass ein breites Spektrum an Methoden zur Erfassung von Emotionen besteht und ebenso verschiedene Theorien und Definitionen. Ziel dieser Arbeit ist vor allem, die subjektive Selbstauskunft der Nutzer zu erfassen, da für die Zufriedenheit vor allem das eigene Empfinden ausschlaggebend ist. Ist der Nutzer subjektiv in einer positiven emotionalen Lage, dann ist unwesentlich, ob ein anderer

Nutzer dies ebenfalls als positiv bewerten würde. Deshalb wird der SAM [Bradley und Lang, 1994] für die Angaben der Nutzer verwendet.

Um aber spontane affektive Reaktionen während der Suche zu erhalten und ein objektives Vergleichsmaß zu generieren, wird zusätzlich ein EMG eingesetzt. Als grundsätzliche Definition von Emotionen wird die eine Bewertung einschließende Reaktion auf kognitive Prozesse einerseits und eine grundsätzlich vorhandene Stimmungslage, die der Nutzer mittels des SAMs angibt andererseits festgelegt.

#### **11.1.5 Methodenauswahl – Nutzerinteraktion**

Die Auswahl der weiteren verwendeten Methoden wird im Folgenden dargelegt.

##### **11.1.5.1 Fragebögen**

Es kamen Fragebögen zum Einsatz, um *explizites* Feedback der Probanden über den Suchverlauf zu erhalten. So kann ein Teil der Variablen, die Bestandteil der Modellzusammenfassung (siehe dazu Kapitel 6.2) sind, erfasst werden.

Vor dem Test, nach dem Test als auch vor und nach jedem Task füllte die Testperson Fragebögen aus. Mit Hilfe dieser Fragebögen wurden demographische Faktoren festgehalten sowie weiter die Erfahrung im Umgang mit der Wikipedia (*pre-study* Fragebogen, siehe Anhang 4.1), die Kenntnis des jeweiligen Themengebiets (*pre-task* Fragebogen, siehe Anhang 4.2) und die subjektive Einschätzung des emotionalen Zustands mit Hilfe des *Self Assessment Manikin* (*post-task* Fragebogen, siehe Anhang 4.3). Außerdem wurden die Testpersonen gebeten, den Erfolg und die Zufriedenheit mit dem Suchverlauf anzugeben (*post-task* Fragebogen). Nach Abschluss der letzten Testaufgabe wurden die Testpersonen um eine Einschätzung der Eignung der Wikipedia als Quelle für die zu erledigenden Aufgaben gebeten und nach allgemeinen positiven oder negativen Auffälligkeiten gefragt (*post-study* Fragebogen, siehe Anhang 4.4). Für alle Ratings in den Fragebögen wurden die sechsstufige Likert-Skala entsprechend dem Wertebereich im Modell, verwendet.

##### **11.1.5.2 Eyetracking Daten**

Um Aufschluss über die von den Nutzern für die Lösung der Suchaufgaben verwendeten Inhaltselemente der deutschsprachigen Wikipedia und seiner Informationsrezeption im Sinne des *impliziten* Feedbacks zu erhalten ist es notwendig, zu protokollieren, wohin der Nutzer auf dem Bildschirm sieht. Ziel dieser Methode ist, die jeweils ver-



brachte Zeit mit einem Inhaltstyp zu erfassen, um Aussagen über die subjektive Relevanz machen zu können (siehe dazu auch [Cutrell und Guan, 2007; Morita und Shinoda, 1994; Agichtein et al., 2006]). Eyetracking eignet sich gut, um die Interessenslage der Nutzer, die Informationsverarbeitungsstrategien und die Reihenfolge der wahrgenommen Inhalte zu untersuchen. Dabei ist vor allem ausschlaggebend, welche Informationselemente Gegenstand von Fixationen sind und darüber hinaus die jeweilige Dauer der Fixation, da die Informationsaufnahme fast ausschließlich nur während der Fixationen stattfindet und nicht während der Augensprünge (Sakkaden) [Joos et al., 2002].

Die Blickbewegungen der Testpersonen wurden deshalb mit Hilfe eines SMI RED EyeTrackers (Software: iView X, Version 2.7.13) aufgenommen. Die dazugehörige *BeGaze* Software liefert auf Basis dieser Aufnahmen Videos von der Interaktion des Nutzers mit dem Browser inklusive der Blickbewegungen auf der Website.

So kann zu jedem Zeitpunkt nachvollzogen werden, wohin der Nutzer auf die Seite gesehen hat. Um die mittels des Eyetrackers erstellten Videos mit genau dieser Information zu annotieren, wurde ein Set von Labels entworfen. Diese Labels beziehen sich erstens auf das gerade vom Nutzer fokussierte Inhaltselement (siehe dazu auch Kapitel 9.5) und zweitens auf die Aktion die er durchführt, wie etwa lesen, navigieren oder ein Bild betrachten (vollständige Tabelle der Labels siehe Tabelle 11-2 und Tabelle 11-3). Die Aktionslabels wurden nach dem typischen Blickverhalten für z.B. Lesen (kurze Sakkaden, Blick folgt einer horizontalen Linie etc.) oder „überfliegen“ (lange vertikale Sakkaden mit kurzen Fixationen, vgl. [Duchowski, 2007]) vergeben.

*Tabelle 11-2 Videolabels "Aktion"*

Aktion	Label-Name	Beschreibung
<b>Lesen</b>	LE	Nutzer liest Text
<b>Querlesen</b>	QL	Nutzer überfliegt Inhalt, z.B. Text, Überschriften, Infoboxen, Listen etc.
<b>Betrachten</b>	BE	Nutzer betrachtet etwas (zum Beispiel ein Bild)
<b>Navigieren</b>	FV	Funktionales Verhalten: Nutzer navigiert

Tabelle 11-3 Videolabels "Inhaltselement"

Label-Name	Inhaltselement
UES	Überschrift
AA	Artikel-Absatz (Text)
LI	Liste (incl. Inhaltsverzeichnis)
EL	Einleitung
BI	Bild (Klassifizierung nach Anhang 1)
IG	Infografik (Klassifizierung nach Anhang 1)
IB	Infobox
IW	Navigation innerhalb der Wikipedia (nur in Kombination mit FV)

Tabelle 11-4 Mögliche Kombinationen der Videolabels

Aktions-Typen	Inhalts-Element-Typen
<b>Lesen [LE]</b>	Artikelabsatz [AA]
	Infografik [IG]
	Liste [LI]
	Bild [BI]
	Infobox [IB]
	Einleitung [EL]
	Links [LI]
<b>Querlesen [QL]</b>	Artikelabsatz [AA]
	Überschrift [UE]
	Infografik [IG]
	Liste [LI]
	Bild [BI]
	Infobox [IB]
	Einleitung [EL]
<b>Betrachten [BE]</b>	Links [LI]
	Infografik [IG]
<b>Funktionales Verhalten [FV]</b>	Bild [BI]
	Innerhalb Wikipedia [IW]
	Browser Aktivität [BA]

### 11.1.5.3 Weitere Methoden

**Browser-Logs:** Mit Hilfe einer zu diesem Zweck implementierten FireFox-Erweiterung wurden bestimmte einfache *events* aus der Interaktion des Nutzers mit dem Browser mitgeloggt, wie z.B. *mousemove*, *click*, *scroll* und *keydown*. Zusätzlich wurde zu jedem Event ein Zeitstempel erfasst.

frame id	mousemove duration	mouseout duration	mouseover duration	mouseup duration	overflow duration	resize duration	scroll duration
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.548	0.0	0.202	0.0	0.0
4	0.0	0.0	1.0	0.0	0.0	0.0	0.0
5	0.0	0.0	1.0	0.0	0.0	0.0	0.0
6	0.654	0.012	0.334	0.0	0.0	0.0	0.0
7	0.994	0.004	0.002	0.0	0.0	0.0	0.0

Abbildung 11-9 Datenbeispiel Browserlogs

**EMG:** Um emotionale Reaktionen des Nutzers während des Tests aufzeichnen zu können, wurde ein Elektromyograph (siehe dazu auch Kapitel 11.1.4.2) verwendet. Dafür wurden fünf runde Silber/Silberchlorid-Elektroden mit einem Durchmesser von 11 Millimetern über der Region zweier Gesichtsmuskeln (dem *musculus zygomaticus major* und dem *musculus corrugator supercilii*) gemäß den Anweisungen von [Cacioppo, 1989] angebracht. Diese Methode sollte ergänzend zur Selbstauskunft mittels SAM [Bradley und Lang, 1994] Aufschluss über affektive Reaktionen der Nutzer geben.

### 11.1.6 Grundlagen zum Aufbau des Experiments

Das Design der empirischen Untersuchungen wurde an der offenen Frage nach Nutzungsszenario-spezifischen Präferenzen bei der Suche in der Wikipedia ausgerichtet.

Das Experiment musste in einem für die Testpersonen akzeptablen zeitlichen Rahmen bleiben, weshalb die Idee verworfen wurde, künstlich Langeweile bei den Probanden zu erzeugen, indem man sie sehr lange warten ließe und in dieser Zeit aber die deutschsprachige Wikipedia zum Zeitvertreib zu Verfügung stellte (als mögliches Szenario für einen *casual leisure task*).

Die Experimente sollten einen zeitlichen Rahmen von maximal sechzig Minuten inklusive Vorbereitungszeit pro Proband nicht überschreiten.

---

Es wurden insgesamt zwei Studien durchgeführt, die ähnlich angelegt waren, wobei beim Design des zweiten Experiments auf die Erfahrungen aus der ersten Erhebung zurückgegriffen und einige Aspekte verbessert werden konnten.

Beide Erhebungen waren im Rahmen von Forschungsseminaren am Lehrstuhl für Informationswissenschaft an der Universität Regensburg angelegt. Design der Experimente, Entwicklung des Versuchsaufbaus im Labor und Erhebung der Daten fand unter Leitung und Aufsicht der Autorin und Prof. Dr. Bernd Ludwig statt, in der ersten Studie zusätzlich mit Dr. David Elweiler.

### 11.1.7 Verwendung des Modells im Experiment

Aufbau der Studien und Wahl der Methoden wurden weiter auch am hier erstellten Modell und damit an der aktuellen Forschungslage ausgerichtet.

Ein wichtiger Schritt war die Präzisierung der Taskbeschreibungen: Die Bezeichnungen *look up*, *learn* und *casual-leisure* werden zur allgemeinen Beschreibung und zum einfacheren Verständnis in den folgenden Auswertungen im Sinne von Task-Klassen verwendet. Nach dem Modell sind die Tasktypen im Experiment wie folgt definiert:

Beispiele (für die vollständige Taskbeschreibung siehe Anhang 3):

Task LU1 :

Am Wochenende wollen Sie mit ihrer Familie einen Ausflug auf die Zugspitze machen. **Um den zeitlichen Ablauf des Tages zu planen** (M = stark *work-based*) wollen Sie in der Wikipedia nachschauen, **wie lange** ( $S_1$  = hoch definierte Information) eine **Fahrt mit der Zahnradbahn dauert** ( $S_2$  = endliches Suchziel, wird genau diese Information gefunden, ist das Suchziel erreicht).

Ein nicht-informationsbezogenes Suchziel wie etwa Zeitvertreib oder Ablenkung ist hier nicht vorhanden.

Die Überführung dieser qualitativen Daten in das quantitative Modell ergibt:

Typ1:

$S_1 = 4-5$	$S_2 = 5$	$S_3 = \text{beliebig, hier } 0-1$	$M = 4-5$
-------------	-----------	------------------------------------	-----------

Task LE3:

Immer wenn Ihre befreundete japanische Austauschstudentin auf die Toilette geht, dreht diese die Musik laut auf, sodass sich eine Nachbarin über die hohe Lautstärke klagt. Da es bereits mehrere Missverständnisse aufgrund kultureller Differenzen zwischen den beiden Damen gegeben hat, **möchten Sie als Vermittler fungieren** (M = stark *work-based*). Hierzu rufen Sie die Wikipedia auf, um sich **über die Toiletten- und Sanitärkultur in Japan zu informieren** ( $S_1$  = Ziel ist relativ klar definiert, nicht mehr ganz so stark wie bei Typ 1) und **ihr Verhalten besser verstehen zu können** ( $S_2$  = ist hier weniger stark ausgeprägt, „besser verstehen“ reduziert die Endlichkeit der Suche).

Ein nicht-informationsbezogenes Suchziel ist wiederum nicht vorhanden.

Typ2:

$S_1 = 3-4$	$S_2 = 2-3$	$S_3 = \text{beliebig, hier eher } 0-1$	$M = 5$
-------------	-------------	---	---------

Die *casual leisure* Aufgaben wurden in den beiden Experimenten unterschiedlich gestellt, jeweils war die Aufgabe, einen bestimmten Zeitraum damit zu verbringen, persönlich interessante Informationen in der Wikipedia zu suchen ohne ein darüber hinausgehendes konkretes Suchziel zu nennen.

Typ3:

$S_1 = \text{beliebig}$	$S_2 = 0-1$	$S_3 > 3$	$M = 0-1$
-------------------------	-------------	-----------	-----------

Durch die Aufgabe im Experiment werden die Werte von  $S_1 - S_3$  und  $M$  vorgegeben. Die Tasks wurden so formuliert, dass eine gewisse Interpretationsfreiheit für den Probanden möglich war [Borlund, 2000], deshalb werden in den Beispielen die Angaben für die Variablen mit einem Spielraum wie 2-3 oder 0-1 gemacht, da Schwankungen durch die Interpretation durch die VPs möglich sind. Generell entsprechen die in der Aufgabenstellung formulierten Motivationen und Ziele den entsprechenden Klassen, (siehe dazu 11.2.2 unten).

Weitere Vorgaben waren durch das Experimentdesign gegeben (in der rechten Spalte befindet sich jeweils die Angabe, wo und wie die Werte der Variablen erfasst werden):

*Tabelle 11-5 im Experiment vorgegebene Variablen*

	Variablenname	Wertebereich	Datenbereich
V <sub>7</sub>	Domain	D <sub>t</sub> = entsprechend der Fragestellung D <sub>CL</sub> = {siehe Klassifikation}	Vorgabe
V <sub>8</sub>	verwendetes Gerät	De= PC	Vorgabe
V <sub>9</sub>	verfügbare Zeit	T <sub>LU</sub> = 2 Min; T <sub>LE</sub> = 4 Min; T <sub>CL</sub> =5 Min	Vorgabe
V <sub>c</sub>	Komplexität	V <sub>c</sub> ( <i>look up</i> ): <i>normal decision task</i> [Byström und Järvelin, 1995] V <sub>c</sub> ( <i>learn</i> ): <i>normal information processing task</i> [Byström und Järvelin, 1995]	Vorgabe

Diese Variablen entsprechen zusätzlich zu M und S<sub>1-3</sub> hier unabhängigen, ursächlichen Variablen. Sie werden durch das Experimentdesign konstant gehalten. Weitere Daten werden während des Experiments erfasst. Die zu erfassenden Variablen sind nicht durchgängig als abhängige Variablen zu betrachten – demographische Angabe sowie Daten zu Vorwissen und emotionaler Selbstauskunft fallen nach der Hypothese dieser Arbeit ebenfalls in die Kategorie der ursächlichen Variablen, die den weiteren Nutzerkontext definieren.



Tabelle 11-6 Im Experiment zu erfassende Daten aus dem Modell

	Variablenname	Wertebereich	Datenbereich
V <sub>4</sub>	Zufriedenheit	Z <sub>1</sub> = [Zufriedenheit mit dem Ergebnis] {0...5} Z <sub>2</sub> = [Zufriedenheit mit dem Verlauf] {0...5}	PT Fr. 5 PT Fr. 4
V <sub>5</sub>	Sucherfolg	E <sub>S1+S2</sub> = [gefundene Information] {0...5} E <sub>S3</sub> = [Erfüllung des nibS] {0...5}	PT Fr.4 PT Fr. 4 (CL)
V <sub>10</sub>	Aktion	$\alpha_t$ = {read, scan, navigate, look at}	Eyetracker FF-Logs
V <sub>12</sub>	betrachtetes Element	E <sub>t</sub> = {text, pic, list, headline, introduction, infobox, chart, links...}	Eyetracker FF-Logs
V <sub>D</sub>	Demographische Faktoren	V <sub>D1</sub> = Jahrgang V <sub>D2</sub> = Geschlecht V <sub>D3</sub> = höchster Bildungsabschluss	PrS Fr. 1-3

Tabelle 11-7 weitere zu erfassende Variablen aus dem Modell

	Variablenname	Wertebereich	Datenbereich
V <sub>14</sub>	Verbleibende Zeit	T <sub>to</sub> = 100% T <sub>t</sub> = {0% - 100%}	Uhr
V <sub>6</sub>	Vorkenntnis	V <sub>Thema</sub> = {0...5} V <sub>WP-Recherche</sub> = {0...5}	PrT Fr. 2 PrS Fr. 5
V <sub>3</sub>	Stimmung	St <sub>t</sub> = {SAM}	PrS Fr.7 PT Fr. 1
V <sub>11</sub>	task completion	C <sub>to</sub> = {0%} C <sub>t</sub> = {0% - 100%}	Ersatzweise: Angaben zum Su- cherfolg

Zu Beginn des Experiments (bzw. bei Start eines Tasks) hat jede Variable einen bestimmten Wert, der das Ausgangsszenario beschreibt. Das Modell erfasst die Werte, die die Variablen zum Zeitpunkt des Auftretens des ASK [Belkin, 1980] oder dem Auftreten eines *nibs* (oder beidem) haben. Es werden konkrete Wertebereiche angegeben. Der Vorteil des Modells liegt darin, dass die Suchanfragen und die Motivation des Nutzers anhand verschiedener Variablen definiert werden kann und es nicht notwen-

dig ist, eine feste Liste an möglichen „Information needs“ festzulegen (wie etwa „*fact retrieval, know item search, verification, navigation, transaction, question answering, knowledge acquisition, comprehension, aggregation, planing, discovery* etc. [Marchionini, 2006]).

Außerdem kann eine bestimmte Zahl ursächlicher Variablen wie  $S_1 - S_3$  oder  $V_8$  und  $V_9$  gleich gehalten werden so dass gezielt andere Einflussgrößen (verwendete Inhalte etc.) untersucht werden können.

Die Kontext des Nutzers sowie sein jeweiliges Suchverhalten lassen sich mit dem Modell zu jedem Zeitpunkt beschreiben.

## 11.2 Erste Datenerhebung

Diese Studie fand im Rahmen des Seminars *Human Activity Recognition* im Sommersemester 2012 an der Universität Regensburg zusammen mit Prof. Dr. Bernd Ludwig und Dr. David Elsweiler statt. Aufbau und Durchführung der Studie geschah gemeinsam mit den Teilnehmern des Seminars, die eigene Forschungsfragen definieren, untersuchen und mit Hilfe der Experimente beantworten sollten.

### 11.2.1 Aufbau der Studie

Die Studie wurde an der Universität Regensburg im Eyetracking-Labor des Lehrstuhls für Informationswissenschaft durchgeführt. Die Versuchspersonen saßen dabei in einem Raum, der Versuchsleiter in einem abgetrennten, durch eine Tür verbundenen Nebenraum. Von dort aus wurde das Experiment gesteuert und überwacht. Auf dem Monitor des Stimulus-PC mit dem Betriebssystem Windows 7 Professional wurden den Testpersonen die Fragebögen, die Aufgabenbeschreibung und die Suchaufgabe selbst präsentiert. Zur Erstellung und Präsentation des Experiments wurde das SMI Experiment Center in der Version 3.2.17 verwendet. Der Versuchsleiter war während des Ablaufs nicht im selben Raum anwesend, um die Testpersonen nicht zu beeinflussen. Für die Bearbeitung der Suchaufgaben stand der Browser Mozilla Firefox zur Verfügung.

Wurde das EMG verwendet (nicht bei allen Versuchspersonen, da eine ausdrückliche, gesonderte Zustimmung dazu notwendig war) musste eine zusätzliche Vorbereitungs-

zeit von ca. 30 Minuten berücksichtigt werde, da das Anbringen der Elektroden zeit-  
aufwändig ist.

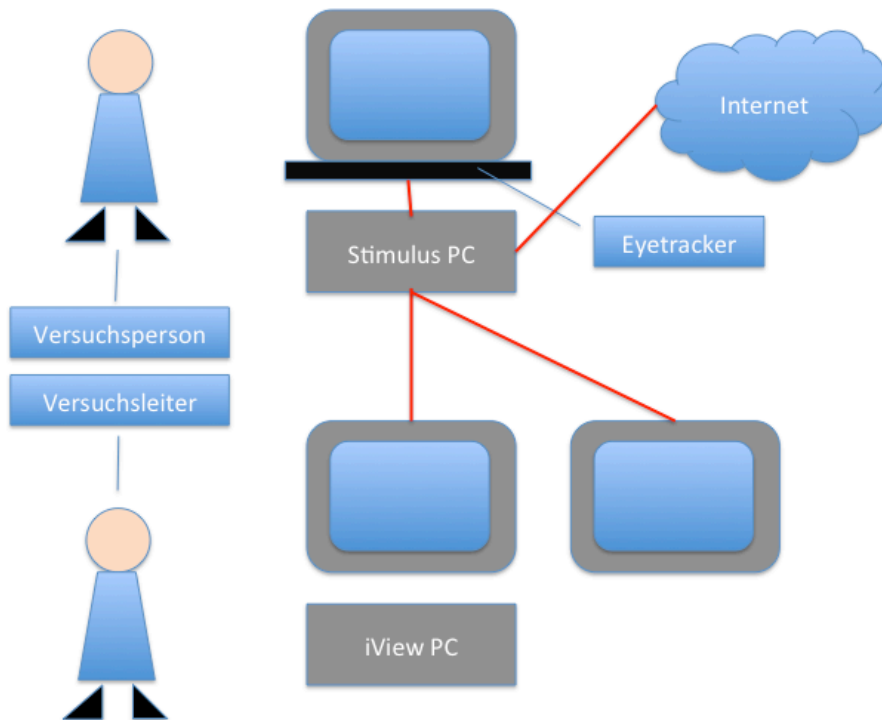


Abbildung 11-10 Laboraufbau Studie I und II

### 11.2.2 Die Aufgaben

Aufgabe der Probanden war es, sechs Suchaufgaben zu lösen. Die Bearbeitungszeit für alle sechs Aufgaben zusammen lag zwischen 30-40 Minuten.

Die Tasks waren so gestaltet, dass sie nach [Marchionini, 2006], [Elsweiler et al., 2011] sowie [Lindley et al., 2012] drei verschiedenen Aufgabentypen entsprachen:

- *Look-up: Fact retrieval, known item search, navigation, transaction, verification, question answering* [Marchionini, 2006]. Entscheidend ist, dass bei dieser Aufgabe eine bestimmte, definierte und endliche Informationseinheit gefunden werden soll, zum Beispiel die Höhe eines Bergs, eine Jahreszahl, ein Name, Öffnungszeiten eines Museums, Abfahrtszeit eine Fähre und dergleichen.

- *Learn: knowledge acquisition, comprehension/interpretation, comparison, aggregation/integration, socialize* [Marchionini, 2006]. Bei dieser Aufgabenart geht es vor allem darum, zu einem bestimmten Zweck eine – mehr oder weniger definierte und zwar theoretisch endliche aber nicht mit definiertem Ende versehene – Menge an Informationen über ein bestimmtes Thema oder einen Sachverhalt zu aggregieren. Zum Beispiel möchte man sich ein Bild über ein bestimmtes Urlaubsziel schaffen, einen Überblick über die Geschehnisse eines Ereignisses wie der Fußball Weltmeisterschaft bekommen oder sich Klarheit zur den Gefahren der Eurokrise verschaffen.
- *Casual Leisure: Non-work Wikipedia tasks* [Elsweiler et al., 2011] und [Lindley et al., 2012] *Non-information focused, hedonistic, serendipitous discovery*. Dabei definieren Elsweiler und Wilson wesentliche Punkte dieses Aufgabentyps: Die Motivation zur Suche entsteht in diesem Fall weniger aus einem konkreten, spezifizierten Informationsbedürfnis sondern zum Beispiel durch den Wunsch, die eigene Stimmung zu verbessern. Der Erfolg eines *casual-leisure* Suchtasks ist dementsprechend auch nicht unbedingt abhängig davon, ob eine bestimmte Information gefunden wurde oder nicht, sondern entsprechend subjektiv (siehe dazu auch *nibs* im Kontextmodell).

### 11.2.3 Formulierung der Aufgaben

Bei der Formulierung der Tasks, die der Nutzer zu Studienzwecken ausführte, waren im Vorfeld einige Punkte zu berücksichtigen. Dazu finden sich Hinweise bei Borlund und Schneider [Borlund und Schneider, 2010] und [Borlund, 2000]. Das von den Autoren dort beschriebene Konzept der *simulated work task situation* [Borlund, 2000] zielt vor allem auf Authentizität und eine sinnvolle Messbarkeit der subjektiven Zufriedenheit der Nutzer mit einem System und den erhaltenen Informationen ab.

In ihrem Konzept zur Evaluierung von *Information Retrieval Interaction* [Borlund, 2000] liefert die Autorin ein Modell, das Vorschläge dazu macht, wie ein Information Retrieval System sinnvoll und realistisch evaluiert werden kann. Dafür konzentrieren sich auch Borlund und Ingwersen [Borlund und Ingwersen, 1997] vor allem auf drei Komponenten:

- 1) „the involvement of potential users as test persons;
- 2) the application of individual and potentially dynamic information need interpretations; and
- 3) the assignment of multidimensional and dynamic relevance judgements“

[Borlund und Ingwersen, 1997, S.19]

Entscheidend für die Anwendung dieser Vorgehensweise ist, dass so der Nutzer selbst in die Entstehung einer Suchaufgabe mit eingebunden wird. Seine persönlichen Faktoren wie die kognitive Struktur und seine emotionale Lage sind damit Teil der Interpretation. Die verschiedenen Dimensionen, die sein Relevanzurteil beeinflussen können werden so berücksichtigt [Borlund und Ingwersen, 1997]. Der Suchtask wird dazu in eine *cover story* eingebettet. [Borlund, 2000] beschreibt zwei Bedingungen, die ein simuliertes Szenario erfüllen muss:

- 1) Es muss eine Situation gegeben sein, die das Bedürfnis nach Information auslöst bzw. ein Bedürfnis schafft, das eine Suche (mit dem Ziel einer veränderten Wissensstruktur) auslöst.
- 2) Die gesuchte Information muss irgendwo vorhanden sein und vom User wahrgenommen werden können, zum Beispiel in einem Informationssystem, mit dem der User interagieren kann.

Bei der Formulierung der Tasks dieser Arbeit ist weiterhin zu berücksichtigen, dass für die Bearbeitung der Aufgaben ausschließlich die deutschsprachige Wikipedia verwendet wird. Die gesuchten Informationen müssen also dort zu finden sein.

Weiterhin stellen Borlund und Ingwersen [Borlund und Ingwersen, 1997, S.81] in Frage, ob es sinnvoll ist, neben der *cover story* überhaupt noch eine zusätzliche, konkrete Fragestellung (*indicative request*) zu präsentieren. Beschreibt man *nur* eine Situation, dann bestimmt der Nutzer das Informationsbedürfnis, das daraus entsteht. Borlund konzentriert sich bei ihren Ausführungen auf *work-based* Szenarios. Deshalb wurde in dieser Arbeit bei der Aufgabenstellung bewusst auf eine ausformulierte, direkt gestell-

te Frage verzichtet, um – trotz Labortest – die Nutzeraktionen so natürlich wie möglich zu halten.

In einem Punkt allerdings unterscheiden sich die hier formulierten Tasks von den Empfehlungen Pia Borlunds (v.a. [Borlund, 2000]), denn es werden zeitliche Restriktion für die Bearbeitung der Aufgaben gegeben. Das ist in diesem Fall notwendig, um erstens eine einigermaßen einheitliche Suchzeit bei allen Probanden zu erhalten und zweitens die Probanden dahingehend etwas „unter Druck“ zu setzen, damit diese sich wirklich auf die von ihnen als wesentlich wahrgenommenen Inhaltselemente konzentrieren. Denn Pre-Tests<sup>45</sup> zeigten, dass *ohne* Zeitlimit die Probanden versucht waren, den entsprechenden Wikipedia-Artikel einfach „von Anfang bis Ende“ durchzulesen. Das entspricht zwar vermutlich auch *einem* von vielen realistischen Nutzungsszenarios, hier war es aber wichtig, dass die Nutzer selektiv vorgehen [Dervin, 1992], um zu erkennen, welche Inhalte konkret bei den verschiedenen Suchaufgaben bevorzugt wurden.

Die *exploratory search* bzw. *casual-leisure* Aufgaben mussten mit dem Ziel, ein von Interesse geleitetes Stöbern auszulösen, formuliert werden. Kules und Capra [Kules und Capra, 2009] beschreiben die Schwierigkeiten bei der Erstellung solcher Aufgaben vor allem darin, dass das *explorative* Vorgehen eigentlich nur ein Ziel auf oberster Eben ist. Die konkreten Gründe für diese Art von Suche sind, wie [Elsweiler et al., 2011; Elsweiler und Wilson, 2010] zeigen, unterschiedlich. Generell ist davon auszugehen, dass eine explorative Suche ein eher unklares Ziel hat und auf der Verfolgung von bestimmte Interessens- und Entdeckungspfaden und/oder dem Serendipity-Effekt basiert [Marchionini, 2006].

Im vorliegenden Versuchsaufbau wird als exploratives Vorgehen ein vom Interesse geleitetes Stöbern in den Inhalten der Wikipedia definiert, dessen erklärtes Ziel es ist, einen vom Versuchsleiter vorgegebenen Zeitraum zu überbrücken.

Durch die randomisierte Reihenfolge der Suchaufgaben (um Lerneffekte zu vermeiden nach [Bortz, 2001]) fiel die Möglichkeit weg, die Versuchspersonen z.B. zu Beginn des Tests einige Minuten „warten“ zu lassen (z.B. unter dem Vorwand, den Eyetracker

---

<sup>45</sup> Das Experimentdesign wurde vor dem Beginn der eigentlichen Datenerhebung getestet.

kalibrieren zu müssen) und ihr dabei zum Zeitvertreib die deutschsprachige Wikipedia anzubieten.

Die Themen und Formulierung der Aufgaben wurden so gewählt, dass sie erstens möglichst ohne Vorwissen verständlich und bearbeitbar waren, zweitens aber nicht so allgemein gestaltet, dass kein authentisches Suchbedürfnis mehr ausgelöst werden kann. Bei der Informationssuche in der Wikipedia sind Aufgaben vorstellbar, die nur anhand mehrerer Artikel gelöst werden können. Da hier auch die Reihenfolge, in der die jeweiligen Artikel studiert werden eine Rolle spielen könnte bzw. der Fokus neben der Auswahl der entsprechenden Inhaltselemente auch auf die Suche nach den entsprechenden Artikeln fallen könnte, wurden für diese Arbeit Aufgabenstellungen gewählt, die anhand *eines* Artikels bearbeitbar waren (nicht gültig für die Aufgaben vom Typ *casual-leisure*).

In beiden Studien gelang es, verständliche Themen von allgemeinem Interesse zu finden. Das zeigte sich vor allem darin, dass während der Tests keine Nachfragen kamen und die Probanden schnell und zielsicher begannen, sich zu informieren. Bei der Erfassung des jeweiligen Interesses und der Vorkenntnis zu den Themen zeigten sich keine extremen Verteilungen (siehe dazu Anhang 5.1), im Mittel war die Vorkenntnis eher niedrig mit einem Wert von 2,14 von max. fünf Punkten, das Interesse an den Aufgabenthemen zeigte einen Mittelwert von 3,19 (n= 103, nur *learn* und *look up* Aufgaben, da bei der freien Suche keine festen Themen vorgegeben waren).

Die Probanden wurden zusätzlich gebeten, zu bewerten, als wie realistisch sie die Aufgabenstellung einschätzten. Hier ergab sich über das gesamte erste Experiment ein Mittelwert von 3,85 (max. fünf Punkte möglich, n = 103). Es wurde auch hier nur bei den *look up* und *learn* Aufgaben nachgefragt, da die Aufgabenstellung für die *casual-leisure* Aufgaben keine Szenario sondern nur eine Anweisung enthielt. Die Grafik in Abbildung 11-11 zeigt, dass die Aufgabenstellungen als relativ realistisch eingeschätzt wurden ( L1, L2 und LE 1 50%- Perzentil bei vier von fünf möglichen Punkten, LE2 bei drei). Realistische Aufgabenstellungen sind die Grundlage für valide Daten.

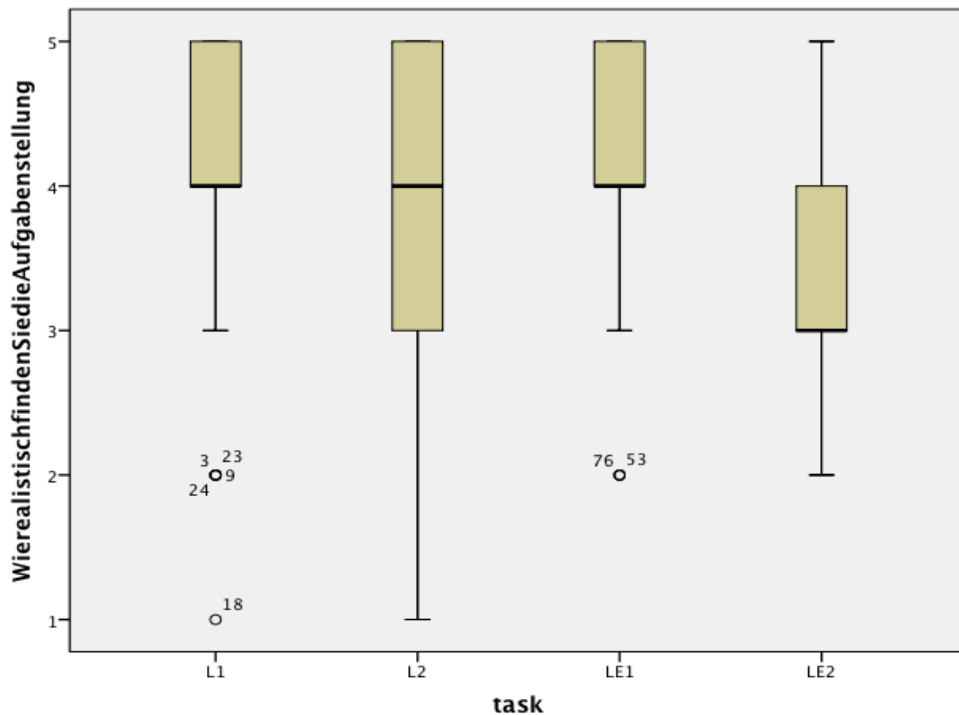


Abbildung 11-11 Angaben der Nutzer, wie realistisch sie die Aufgabenstellung einschätzten

### 11.3 Auswertung der Daten

Die erfassten Daten wurden hinsichtlich der offenen Fragestellung nach den Einflüssen der Nutzerszenarien bei der Informationssuche in der deutschsprachigen Wikipedia ausgewertet.

Dazu wurden verschiedene statistische Verfahren verwendet. Zunächst wurden die Angaben der Probanden zu Vorkenntnis und Nutzung der Wikipedia sowie ihre demographischen Angaben ausgewertet. Es folgt eine Verarbeitung der Eyetracking Videos und die Analyse des Rezeptionsverhaltens der Probanden bei der Lösung der Aufgaben. Dazu wurde die Verteilung der Nutzeraktionen bei den verschiedenen Aufgabentypen untersucht.

Der H-Test von Kruskal und Wallis kam dabei als Verfahren zum Vergleich der zentralen Tendenz von mindestens ordinal skalierten Variablen in mehr als zwei Stichproben als OMNIBUS-Test zum Einsatz [Leonhart, 2013, S. 249]. Das Erkenntnisziel war hier, ob die Nutzeraktionen auf die drei Aufgabentypen unterschiedlich verteilt sind und ob



auch auf eine unterschiedliche Verteilung in der Grundgesamtheit geschlossen werden kann.

Alle bei den jeweiligen Auswertungen berücksichtigten Variablen wurden auf Normalverteilung (Vergleich mit der bei den jeweils vorliegenden Lageparametern entstehenden Normalverteilung) und auf Varianzhomogenität zwischen den Stichproben geprüft, um so die entsprechend passenden statistischen Verfahren auswählen zu können. Das alpha-Fehler-Niveau wurde festgelegt und bei den verschiedenen Tests nach Bonferoni [Leonhart, 2013] korrigiert. Zur Veranschaulichung der Daten und mancher Ergebnisse wurden zusätzliche graphische Darstellungen eingesetzt.

Wurden die zu vergleichenden Stichproben anders als den drei Aufgabentypen entsprechend eingeteilt, sind jeweils bei der Einteilung Hinweise auf die Gründe angegeben (dies gilt v.a. für stetige Variablen aus dem Modell in 6.2).

Die Ergebnisse wurden nach Abschluss beider Studien auf das Modell übertragen.

### 11.3.1 Demographische Angaben

An der ersten Studie nahmen 26 Personen teil, davon waren 11 männlich und 15 weiblich. Die Personen waren zum Zeitpunkt der Erhebungen zwischen 18 und 27 Jahre alt, der Altersdurchschnitt betrug 22,7 Jahre. Jeder Proband erledigte sechs Suchaufgaben. So ergab sich eine Summe von 156 ausgeführten Tasks. Bis auf zwei Personen waren alle Probanden Studierende an einer Hochschule. Die Teilnehmer schätzten ihre Erfahrung mit Wikipedia-Recherchen bis auf eine Ausnahme als gut bis sehr gut ein. Die Probanden gaben außerdem mit durchschnittlich 4,17 Punkten (0-5 Punkte möglich, siehe dazu *Pre-Study* Fragebogen im Anhang 4.1) an, dass sie die Wikipedia als Informationsquelle benutzen. Hier bestätigte sich der eingangs genannte hohe Bekanntheits- und Verwendungsgrad der Wikipedia.

Die Probanden ( $n = 26$ ) wurden weiter gebeten, mit 0 bis 5 Punkten zu bewerten, ob sie die Wikipedia sowohl in *work-based* als auch in Freizeitszenarien verwenden (siehe Abbildung 11-12). Es zeigte sich, dass beide Fälle vertreten sind, allerdings die *work-based* Szenarien überwiegen (Durchschnittlich 4,12 von fünf Punkten, Freizeitszenarien durchschnittlich 3,08 von fünf Punkten, siehe Abbildung 11-12).

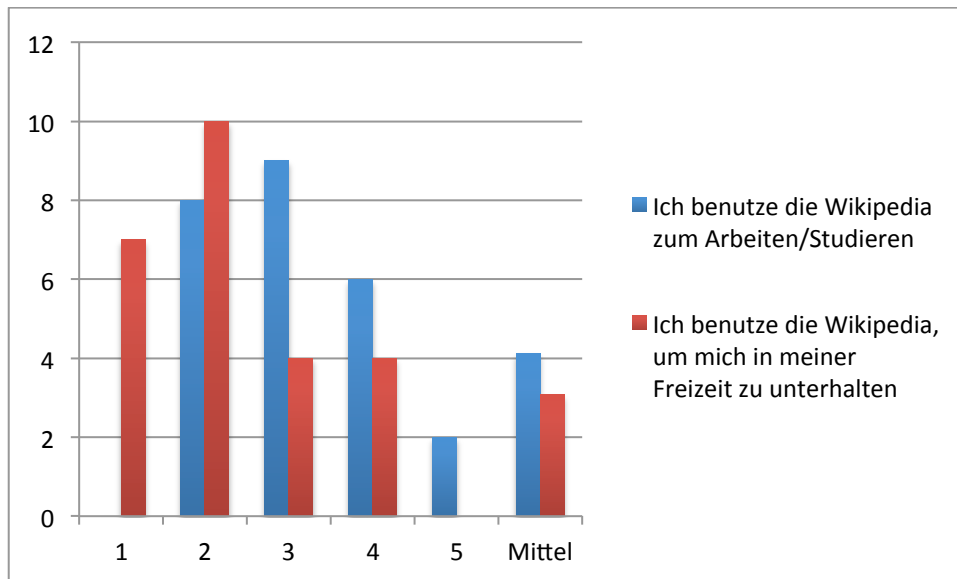


Abbildung 11-12 Nutzung der Wikipedia bei den Probanden

Die Datenerhebung erfolgte ca. von April – Juni 2012 an der Universität Regensburg.

### 11.3.2 Verarbeitung der Videos des Eyetrackers

Die mit Hilfe des Eyetrackers erstellten Videos wurden mit den speziell dafür entworfenen Labels annotiert. Zu jedem Zeitpunkt erhielten die Videos ein Label betreffend das Inhaltselement, das in diesem Moment betrachtet wurde (zum Beispiel ein Textabschnitt) und die Aktion, die der Leser ausführte (zum Beispiel Lesen). Zur Annotation wurde die freie Software „Anvil“ verwendet.

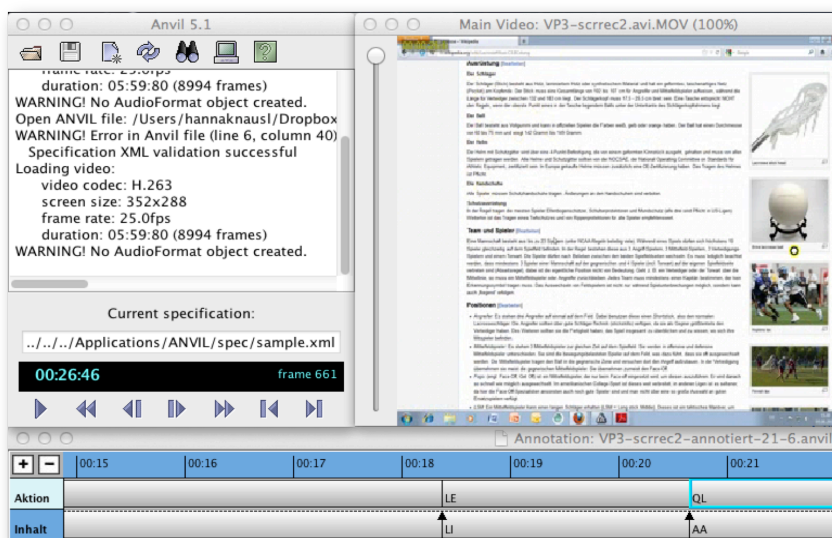


Abbildung 11-13 Software "Anvil" zur Annotation der Videos

Die Videolabels wurden als .txt Datei extrahiert. Diese Dateien enthalten jeweils Start- und Endzeitpunkt der Labels (Beispiel: Siehe Anhang 5.2). Für die weitere Verarbeitung wurden diese Zeitspannen der Labels in 50-ms-Frames unterteilt. So entstand für jeden Suchtask ein Datensatz, der aus  $n$  Frames à 50 ms besteht, von denen jeder eindeutig durch zwei Labels gekennzeichnet ist.

Ziel der Analyse der Videos ist es, Hinweise auf unterschiedliche Häufigkeiten der Videolabels (die die Nutzeraktionen bzw. Informationsrezeption repräsentieren) in den unterschiedlichen Suchaufgaben zu finden.

Nielsen [Nielsen, 1989] bemerkt zu diesem Punkt, dass die *größten* Unterschiede im Verhalten bei der Suche vor allem auf individuellen Unterschieden zwischen den Nutzern und der Art des Tasks basieren [Nielsen, 1989]. Deshalb wird hier weiterhin vor allem der *Tasktyp* berücksichtigt<sup>46</sup>, wobei dieser hier genauer spezifiziert ist als bei [Nielsen, 1989] und konkrete Auswirkungen auf die Verwendung der Inhaltselemente in der deutschsprachigen Wikipedia untersucht werden.

Zuerst wird die Verteilung der Label-Kombinationen zwischen den verschiedenen Tasktypen ausgewertet. Gezählt wird dabei die Häufigkeit der 50-ms-Frames einer Nutzeraktion pro Task. Für die Auswertungen verwendbar waren hier letztendlich insgesamt 90 Suchtasks, je 30 pro Klasse/Typ. Die anderen fielen durch verschiedene Probleme aus der Analyse (z.B. beschädigte oder nicht vorhandene Videodateien).

### 11.3.3 Kruskal-Wallis-Test zur Überprüfung von Unterschieden

Ein Kruskal-Wallis-Test für unabhängige Stichproben ist ein nichtparametrischer Test zur Überprüfung, ob die zentrale Tendenz von mehr als zwei verschiedenen Stichproben unterschiedlich ist. Der Test gibt auch Auskunft darüber, wie groß die Wahrscheinlichkeit ist, dass dieser Trend auch in der Grundgesamtheit zu finden ist.

Genau wie der Whitney-Mann-U-Test ist er rangplatzbasiert, allerdings gibt es bei den vorliegenden Daten *drei* unabhängige Stichproben (jede Gruppe entspricht einem Aufgabentyp), weshalb der Kruskal-Wallis-Test zum Einsatz kommt. Dieser Test setzt keine Normalverteilung der Daten voraus, allerdings müssen diese mindestens ordinalskaliert sein [Leonhart, 2013]. Deshalb werden die Datensätze der Tasks dementsprechend aufbereitet.

---

<sup>46</sup> Als eine von mehreren Einflussvariablen auf das Nutzerverhalten.

Zunächst erhält man für jeden 50-ms-Frame eine Angabe, welche Aktion der Nutzer zu diesem Zeitpunkt durchführt (zu den Abkürzungen siehe Tabelle der Videolabels):

Frame ID	VP8 LE2	VP8 LE1	VP7 LE2	VP7 LE1	VP6 LE2	VP6 LE1	VP3 L1	VP3 L2
	mode duration	mode duration	mode duration	mode duration	mode duration	mode duration	mode duration	mode duration
1	FV-IW <sup>47</sup>	FV-IW	FV-BA	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
2	FV-IW	QL-LI	FV-BA	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
3	FV-IW	QL-LI	FV-BA	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
4	FV-IW	QL-LI	FV-BA	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
5	FV-IW	QL-LI	FV-BA	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
6	FV-IW	QL-LI	FV-BA	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
8	FV-IW	QL-LI	BE-BI	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
9	FV-IW	FV-IW	BE-BI	FV-BA	LE-EL	FV-IW	FV-IW	FV-IW
9	FV-IW	FV-IW	BE-BI	FV-BA	LE-EL	FV-IW	LE-EL	LE-EL
10	FV-IW	FV-IW	BE-BI	QL-EL	LE-EL	FV-IW	LE-EL	LE-EL
11	FV-IW	FV-IW	BE-BI	QL-EL	LE-EL	QL-LI	LE-EL	LE-EL
12	FV-IW	FV-IW	BE-BI	QL-EL	LE-EL	QL-LI	LE-EL	LE-EL
13	FV-IW	FV-IW	BE-BI	QL-EL	LE-EL	QL-LI	LE-EL	LE-EL
14	FV-IW	FV-IW	LE-LI	BE-BI	LE-EL	QL-EL	LE-EL	LE-EL
15	FV-IW	FV-IW	LE-LI	QL-EL	LE-EL	QL-EL	LE-EL	LE-EL

Tabelle 11-8 Datenbeispiel Videolabels/Nutzeraktionen

Im nächsten Schritt wurden die jeweiligen *Häufigkeiten* der Nutzeraktionen *pro Task* ermittelt. Für jede Aufgabe entstand so ein Datensatz, der metrische Angaben für die Häufigkeit des Auftretens jeder Nutzeraktion enthält.

Tabelle 11-9 Datenbeispiel Häufigkeiten Nutzeraktionen/Task

	BE-BI	BE-IB	BE-IG	LE-AA	LE-EL	LE-IB
VP8 LE2	31	0	5	316	147	48
VP8 LE1	23	0	29	340	0	0
VP7 LE2	82	0	0	461	0	0
VP7 LE1	53	0	0	184	64	62
VP6 LE2	31	0	0	206	25	0
VP6 LE1	5	0	0	138	6	0
VP3 L1	18	0	46	478	14	0
VP3 L2	18	0	46	478	14	0
VP2 LE1	32	0	32	397	0	0

<sup>47</sup> Für die Abkürzungen der Videolabels siehe Kapitel 11.1.5.2.

Für einen visuellen Überblick wurden die Mittelwerte der absoluten Häufigkeiten des Auftretens der Videolabelkombinationen aller drei Aufgabentypen graphisch dargestellt (für die genauen Werte und Lageparameter der einzelnen Variablen siehe Anhang 5.3).

n (50 ms frames)

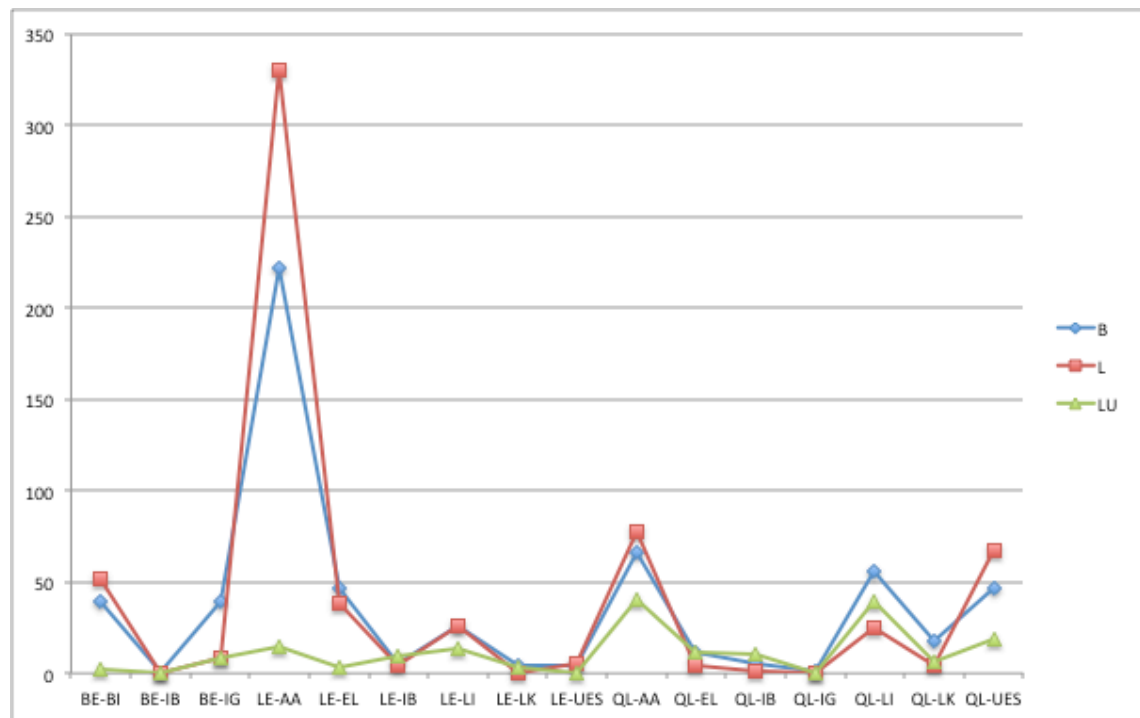


Abbildung 11-14 Mittelwerte der Häufigkeiten der Videolabel-Kombinationen (B= casual leisure, L = Learn, LU = look up).

Die Grafik der Mittelwerte der Häufigkeiten der Videolabel-Kombinationen zeigt Unterschiede im Verlauf. Auffällig bei Abbildung 11-14 ist vor allem, dass alle Labels, die die Informationsaufnahme betreffen (BE-BI bis LE-LK<sup>48</sup>) bei den *look up* Tasks unter den Werten der anderen beiden Aufgaben liegen, wogegen die Häufigkeiten bei den *orientierenden* Labels (QL-EL bis QL-UES) teilweise darüber oder gleich sind.

Nachfolgende Grafik (Abbildung 11-15) stellt die unterschiedliche Zusammensetzung des Suchverhaltens (hier in Form der Mittelwerte der absoluten Häufigkeiten über alle Testpersonen) bei den verschiedenen Aufgabentypen dar. Für diese Darstellung wurden die durch den Kruskal-Wallis-Test (siehe Tabelle 11-10) entdeckten unterschiedlich

<sup>48</sup> Für die Abkürzungen der Videolabels siehe Kapitel 11.1.5.2.

verteilten Videolabelkombinationen verwendet. Besonders fällt die Dominanz der anteiligen Lesezeit (hier rot dargestellt) bei *learn* und *casual-leisure* Aufgaben auf. Der große hellblaue Bereich beim Tasktyp *look up* zeigt, dass ein wesentlicher Teil der Aufgabenzeit mit der Berücksichtigung des Inhaltsverzeichnisses (QL-LI) verbracht wurde (Streckung der Grafik auf 100%). Auch die im Vergleich zum vollständigen Artikel relativ kurze Einleitung wurde vor allem bei *casual-leisure* stark berücksichtigt. Für ein exploratives Vorgehen ist dementsprechend die zusammenfassende Information über den Inhalt des Artikels wichtig.

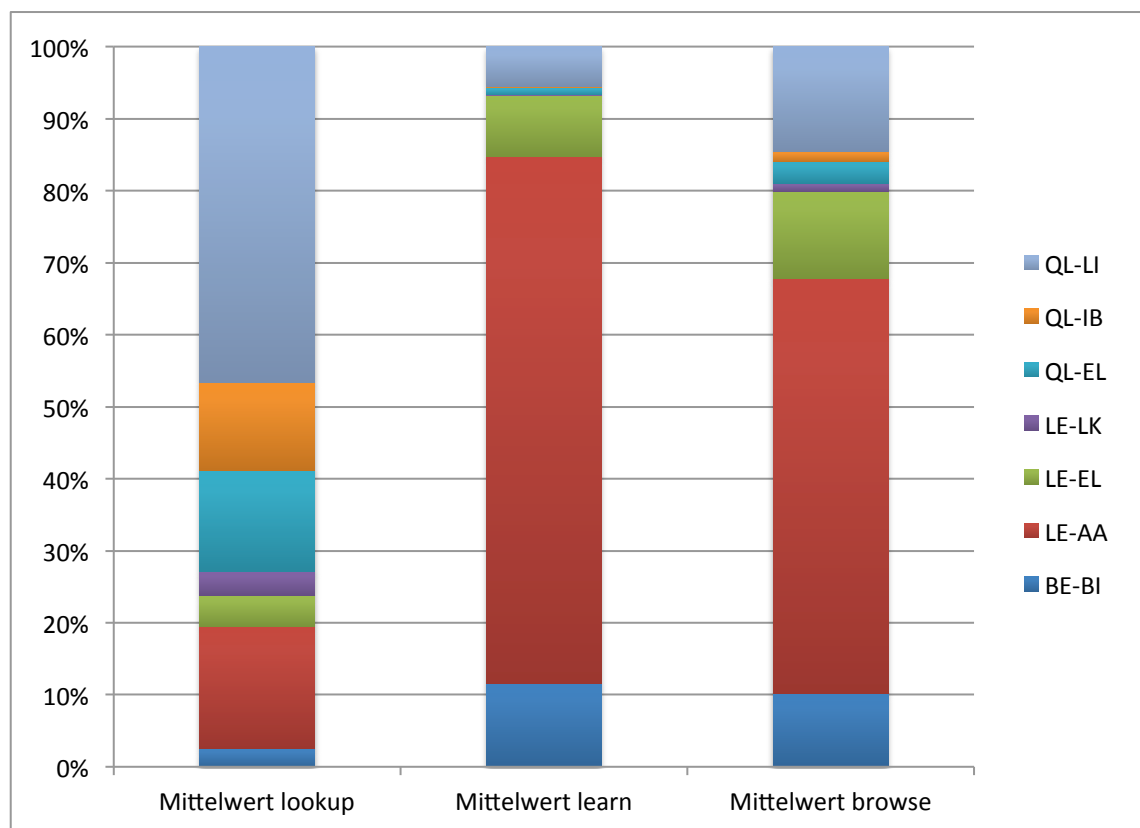


Abbildung 11-15 Aufsummierung aller nach Kruskal-Wallis-Test statistisch signifikant unterschiedlich verteilten Kombinationen (Mittelwerte, absolute Häufigkeiten)<sup>49</sup>

Die Datensätze der jeweiligen Aufgabengruppen bilden im nächsten Test je eine Stichprobe. Der Kruskal-Wallis-Test wird angewendet, um die Gleichverteilung der Videolabel-Kombinationen (Nutzeraktionen) über die drei Aufgabentypen zu untersuchen ( $n_{\text{Aufgabe}} = 30$ ,  $n_{\text{gesamt}} = 90$ ,  $n_{\text{getestete Variablen}} = 16$ , Skalierung der Variablen metrisch, abhängige Variable ist der Aufgabentyp, nominal skaliert).

<sup>49</sup> Für die Abkürzungen der Videolabels siehe Kapitel 11.1.5.2.

Die Ergebnisse des Tests zeigen, dass die Label BE-BI (Bilder betrachten), LE-AA (Artikel-Absatz lesen), LE-EL (Einleitung lesen), BE-IG (Infografik betrachten), QL-EL (Einleitung querlesen), QL-LK (Link querlesen), LE-UES (Überschriften lesen), QL-UES (Überschriften querlesen) und QL-LI (Querlesen einer Liste. In der Regel ist hier damit das Inhaltsverzeichnis gemeint. Durch die Labels kann das nicht dargestellt werden, diese Schwäche wurde bei den überarbeiteten Labels in der Folgestudie behoben. In der ersten Studie musste diese Information den Videos direkt entnommen werden, da die Labels dahingehend nicht aussagekräftig genug waren) über die Tasktypen nicht gleichverteilt sind (siehe Tabelle 11-10). Der Test gibt weiterhin die Wahrscheinlichkeit für einen Fehler erster Art an (hier von 0,00% bis 2,2%). Für statistische Untersuchungen muss eine maximale Wahrscheinlichkeit für einen Fehler erster Art festgelegt werden, die noch akzeptiert wird.

Für die Festlegung des alpha-Fehler Niveaus sind vor allem *Konsequenzen* eines Fehlers erster Art zu berücksichtigen. Im vorliegenden Fall ist die Konsequenz, dass ein Inhaltstyp fälschlicher Weise für eine Informationspräsentation gewählt wird bzw. dass eine Aufgabe, die anhand der Nutzeraktion klassifiziert wird, der falschen Gruppe zugeordnet wird.

Da diese Konsequenzen nicht dramatisch sind (im Gegensatz z.B. zu medizinischen Studien) und vor allem weil die Analyse der Daten explorativ erfolgt, wird das alpha-Fehler Niveau auf eher hohe 10% festgelegt (siehe dazu auch [Leonhart, 2013 S. 181-185]). Die Irrtumswahrscheinlichkeit für eine Zurückweisung von  $H_0$  darf damit nicht über 10% liegen. Wird mehr als eine Variable getestet wird das Fehlerniveau jeweils nach Bonferoni korrigiert [Leonhart, 2013], also auf alle Variablen aufgeteilt ( $10\%/n$ ).

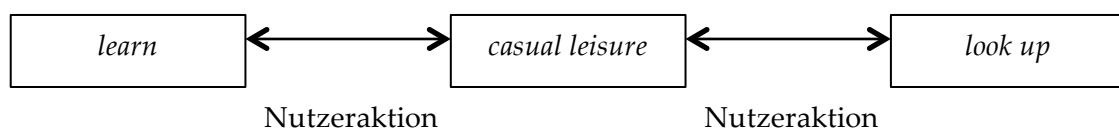


Abbildung 11-16 Modell für den Kruskal Wallis Test

Tabelle 11-10 Ergebnisse des Kruskal Wallis Test über alle drei TT, absolute Häufigkeiten

Übersicht über Hypothesentest				
	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von BE-BI ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
2	Die Verteilung von BE-IB ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,421	Nullhypothese behalten.
3	Die Verteilung von BE-IG ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,004	Nullhypothese ablehnen.
4	Die Verteilung von LE-AA ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
5	Die Verteilung von LE-EL ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
6	Die Verteilung von LE-IB ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,565	Nullhypothese behalten.
7	Die Verteilung von LE-LI ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,463	Nullhypothese behalten.
8	Die Verteilung von LE-LK ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,015	Nullhypothese ablehnen.
9	Die Verteilung von LE-UES ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,016	Nullhypothese ablehnen.
10	Die Verteilung von QL-AA ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,080	Nullhypothese behalten.
11	Die Verteilung von QL-EL ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,069	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

Übersicht über Hypothesentest				
	Nullhypothese	Test	Sig.	Entscheidung
12	Die Verteilung von QL-IB ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,003	Nullhypothese ablehnen.
13	Die Verteilung von QL-IG ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,132	Nullhypothese behalten.
14	Die Verteilung von QL-LI ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,022	Nullhypothese ablehnen.
15	Die Verteilung von QL-LK ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,058	Nullhypothese behalten.
16	Die Verteilung von QL-UES ist über Kategorien von TT (1= Browse, 2 = LE, 3 = L) gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

### 11.3.4 Vergleich der Mittelwerte nach Aufgabentypen

Um weiter zu analysieren, zwischen *welchen* Gruppen die in 11.3.3 ermittelten statistisch signifikanten Unterschiede zwischen den drei Aufgabentypen zu finden sind, wurde als post-hoc Test ein Mehrfachvergleich (im Rahmen einer Varianzanalyse) durchgeführt.

Damit wird überprüft, ob eine statistisch signifikante Mittelwertdifferenz der Häufigkeiten der Nutzeraktionen zwischen den drei Typen bzw. zwischen je zwei der drei angenommen werden kann. Da noch keine konkreten Hinweise vorliegen, welche Stichproben sich in welchen Punkten wie unterscheiden, wird bei der Analyse der Daten explorativ vorgegangen (post-hoc-Test) und kein hypothesengeleitetes Verfahren gewählt. Die Frage nach einer möglichen Mittelwertdifferenz zwischen den Stichproben ist demnach ungerichtet,  $H_0$  ist, dass die Mittelwertdifferenz um null liegt.



Um das entsprechende Verfahren auszuwählen zu können wird die Verteilung der Videolabels in den einzelnen Gruppen untersucht. Mittels Kolmogorov-Smirnov- bzw. Shapiro-Wilk-Test wird überprüft, ob die Hypothese, dass die Nutzeraktionen in der Grundgesamtheit normalverteilt sind, aus den Werten der Stichprobe plausibel herzu-leiten ist [Leonhart, 2013]. Mit dem Levene-Test wurde weiterhin die Varianzgleichheit zwischen den Gruppen untersucht, da auch die Varianzhomogenität Bedingung für einige statistische Verfahren ist.

Da die Verteilung der Videolabels in den Gruppen nicht der Normalverteilung entspricht (siehe Anhang 5.4) und keine Varianzhomogenität vorliegt (siehe Anhang 5.5) wurde als post-hoc-Test der Games-Howell Mehrfachvergleich gewählt, da dieser robust gegen diese Bedingungen ist [Leonhart, 2013].

Die Gruppen entsprechen dabei den Tasktypen *casual-leisure* (B), *learn* (LE) und *look up* (L). Ziel ist es, zu untersuchen, ob die beobachteten Unterschiede der Mittelwerte der einzelnen Gruppen ausreichend groß sind, um davon auf Unterschiede in den zugehörigen Grundgesamtheiten schließen zu können und herauszufinden, zwischen welchen der drei Gruppen die Unterschiede liegen. Es wurden die absoluten Häufigkeiten der 50 ms Frames der nach dem Kruskal-Wallis-Test unterschiedlich auf die drei Gruppen verteilten Videolabelkombinationen verwendet.

Tabelle 11-11 Games - Howell Mehrfachvergleich, absolute Häufigkeiten

Mehrfachvergleiche							
Games-Howell							
Abhängige Variable	(I) V1	(J) V1	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
BE-BI	1	2	-12,700	9,601	,388	-35,80	10,40
		3	37,167*	7,103	,000	19,64	54,69
	2	1	12,700	9,601	,388	-10,40	35,80
		3	49,867*	6,520	,000	33,78	65,95
	3	1	-37,167*	7,103	,000	-54,69	-19,64
		2	-49,867*	6,520	,000	-65,95	-33,78
BE-IG	1	2	30,733	12,581	,052	-,19	61,65
		3	30,433	12,919	,061	-1,18	62,04
	2	1	-30,733	12,581	,052	-61,65	,19
		3	-,300	4,879	,998	-12,08	11,48
	3	1	-30,433	12,919	,061	-62,04	1,18
		2	,300	4,879	,998	-11,48	12,08
LE-AA	1	2	-108,833*	41,046	,028	-207,67	-10,00
		3	207,300*	32,116	,000	128,11	286,49
	2	1	108,833*	41,046	,028	10,00	207,67

	3	3	316,133*	26,148	,000	251,71	380,56
		1	-207,300*	32,116	,000	-286,49	-128,11
		2	-316,133*	26,148	,000	-380,56	-251,71
LE-EL	1	2	8,067	11,010	,745	-18,42	34,56
		3	42,733*	7,588	,000	24,11	61,36
	2	1	-8,067	11,010	,745	-34,56	18,42
		3	34,667*	8,419	,001	13,98	55,35
	3	1	-42,733*	7,588	,000	-61,36	-24,11
		2	-34,667*	8,419	,001	-55,35	-13,98
LE-LK	1	2	4,000	1,998	,128	-,92	8,92
		3	1,633	2,931	,843	-5,42	8,69
	2	1	-4,000	1,998	,128	-8,92	,92
		3	-2,367	2,218	,541	-7,83	3,09
	3	1	-1,633	2,931	,843	-8,69	5,42
		2	2,367	2,218	,541	-3,09	7,83
LE-UES	1	2	-1,467	2,608	,841	-7,80	4,86
		3	3,733*	1,201	,011	,78	6,69
	2	1	1,467	2,608	,841	-4,86	7,80
		3	5,200	2,338	,084	-,57	10,97
	3	1	-3,733*	1,201	,011	-6,69	-,78
		2	-5,200	2,338	,084	-10,97	,57
QL-IB	1	2	5,133*	1,954	,034	,33	9,94
		3	-4,533	3,719	,448	-13,53	4,46
	2	1	-5,133*	1,954	,034	-9,94	-,33
		3	-9,667*	3,214	,014	-17,59	-1,74
	3	1	4,533	3,719	,448	-4,46	13,53
		2	9,667*	3,214	,014	1,74	17,59
QL-LI	1	2	30,833*	11,013	,021	4,08	57,59
		3	16,200	11,632	,353	-11,93	44,33
	2	1	-30,833*	11,013	,021	-57,59	-4,08
		3	-14,633	7,763	,153	-33,33	4,07
	3	1	-16,200	11,632	,353	-44,33	11,93
		2	14,633	7,763	,153	-4,07	33,33
QL-UES	1	2	-20,233	13,864	,318	-53,62	13,15
		3	28,033*	10,338	,025	3,04	53,03
	2	1	20,233	13,864	,318	-13,15	53,62
		3	48,267*	12,048	,001	19,02	77,51
	3	1	-28,033*	10,338	,025	-53,03	-3,04
		2	-48,267*	12,048	,001	-77,51	-19,02

\*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

Da neun abhängige Variablen gleichzeitig getestet werden, muss das alpha-Fehler Niveau dementsprechend korrigiert werden. Die eingangs festgelegte maximale Irrtumswahrscheinlichkeit von 10% wird auf die neun Tests verteilt. Damit wird das alpha-Fehler Niveau hier pro Variable auf  $10/9$ , also eine maximale Irrtumswahrscheinlichkeit von 1,11 %, festgelegt.

Damit wird allerdings keine Aussage über die Stärke oder Bedeutung eines Zusammenhangs gemacht. Es wird lediglich eine Obergrenze für die Irrtumswahrscheinlichkeit der Annahme, die Zusammenhänge seien nicht zufällig, gegeben.

Der Test zeigt folgende Ergebnisse: Die beiden Tasktypen *casual-leisure* und *learn* (in der Tabelle 11-11 codiert als 1 und 2) unterscheiden sich in der Differenz der Mittelwerte der absoluten Häufigkeiten vor allem bei der Zeit, die mit der Lektüre der Texte verbracht wurde.

Zwischen *casual-leisure* und *look up* (in der Tabelle codiert als 1 und 3) dagegen gibt der Test für die Videolabel-Kombinationen BE-BI (Bilder betrachten), LE-AA (Text lesen), LE-EL (Einleitung lesen), an, dass mit einer Irrtumswahrscheinlichkeit von unter oder gleich 1,11 % darauf geschlossen werden kann, dass die Unterschiede (mittlere Differenz der Mittelwerte) zwischen den beiden Gruppen nicht zufällig sind. Ähnlich beim Vergleich zwischen *learn* und *look up* (in der Tabelle codiert als 2 und 3): Bei BE-BI, LE-AA, LE-EL und QL-UES unterscheiden sie sich statistisch signifikant in den Mittelwerten.

Zwischen *casual-leisure/learn* und *look up* Aufgaben unterscheiden sich die Mittelwerte der absoluten Häufigkeiten statistisch signifikant bei den entsprechenden Videolabel-Kombinationen, bei *casual-leisure* vs. *look up* ab noch zusätzlich beim Lesen der Überschriften. Die Videolabel-Kombinationen BE-BI (Bilder betrachten) und LE-AA (Lektüre von Text) betreffen die Informationsaufnahme (im Gegensatz zu orientieren und navigieren).

Die Überschriften oder die Einleitung zu lesen dient dazu, einen Überblick über die jeweilige Thematik zu bekommen. Es wird also abhängig vom Nutzungsszenario unterschiedlich viel gelesen und es werden unterschiedlich oft Bilder angesehen. Auch die Einleitung wird im Mittel nicht vergleichbar oft gelesen, dasselbe gilt für die Überschriften.

Das weist darauf hin, dass unterschiedliche Suchstrategien angewendet werden und die Relevanz der Inhaltselemente für den Nutzer abhängig von seiner Aufgabe entsteht. Dabei werden Unterschiede vor allem bei der Funktion der Inhaltselemente gemacht (siehe dazu auch Grafik in Abbildung 11-14), je nach Aufgabentyp werden zu-

sammenfassende und überblicksartige Artikel-Teile (wie etwa Überschriften und Inhaltsverzeichnis) oder ausführliche (Artikel-Text) und möglicherweise auch „nur“ dekorative (z.B. Bilder) Elemente berücksichtigt.

Ein interessanter Aspekt ist hierbei, dass sich deutlichere Zusammenhänge zeigten, als die Daten zunächst nur für eine kleine Zahl der Probanden ausgewertet wurden, um erste Trends zu erkennen (siehe [Knäusl und Ludwig, 2013], [Knäusl et al., 2012]).

*Tabelle 11-12 Chi-Quadrat Test über die Verteilung der Nutzeraktionen in den drei Tasktypen bei sechs VPs, absolute Häufigkeiten [Knäusl et al., 2012]*

action	LO vs RE		LO vs CA		RE vs CA	
	$\chi^2$	p-value	$\chi^2$	p-value	$\chi^2$	p-value
EX	9	0.011	9	0.029	18	0.006
NV	9	0.011	9	0.011	18	0.001
RE	13	0.043	6	0.301	27	0.079
SC	36.563	0.064	45	0.039	45	0.039

Je mehr Datensätze in die Analyse mit einbezogen wurden, desto geringer wurden die statistisch signifikanten Unterschiede der Ausprägung der einzelnen Zustände zwischen den Tasktypen. Entsprechend [Nielsen, 1989] entstand daraus die Vermutung, dass die Nutzer auch grundsätzlich (nicht *nur* durch Kontextfaktoren und hier vor allem dem Aufgabentyp beeinflusst) unterschiedlich suchen. Normierung der Häufigkeiten der Labelkombinationen

Um das Problem der unterschiedlichen Videolängen<sup>50</sup> zu beheben wurde im Weiteren zusätzlich die jeweils relative Häufigkeit der Videolabel-Kombinationen im Task untersucht. Dazu wurden jeweils zwei gleiche Suchaufgaben pro Proband zur Berechnung herangezogen, über alle Probanden erhält man so eine Wahrscheinlichkeit, mit der die jeweilige Nutzeraktion bei den einzelnen Aufgabentypen zu erwarten ist. Damit wird die Häufigkeit der Nutzeraktionen an der Summe der Interaktionen normiert.

Für jede VP wird für jeden Tasktyp die relative Häufigkeit für die einzelnen Videolabel-Kombinationen  $h_n(VK_x)$  berechnet mit der Formel:

<sup>50</sup> Verursacht entweder durch fehlerhafte Dateien oder unterschiedlich lange Bearbeitungszeit der einzelnen Aufgaben.

## Formel 1 Berechnung der relativen Häufigkeiten der Videolabelkombinationen

$$h_n(VK_x) = \left( \frac{n_{Vx\ T1}}{n_{Frames\ T1}} \right) * \left( \frac{n_{Frames\ T1}}{n_{Frames\ T1} + n_{Frames\ T2}} \right) + \left( \frac{n_{Vx\ T2}}{n_{Frames\ T2}} \right) * \left( \frac{n_{Frames\ T2}}{n_{Frames\ T1} + n_{Frames\ T2}} \right)$$

## 11.3.5 Verteilung der Wahrscheinlichkeiten

Es wurde die Verteilung der Wahrscheinlichkeiten  $p(\text{Videolabel-Kombination})$ <sup>51</sup> bei den drei Aufgabentypen untersucht.

Die graphische Darstellung der Mittelwerte von  $h_n$  in Abbildung 11-17 zeigt zunächst, dass die Verteilung der Wahrscheinlichkeiten für die Videolabels teilweise ähnlich ist, teilweise aber auch voneinander abweicht.

$p(\text{Videolabel-Kombination})$

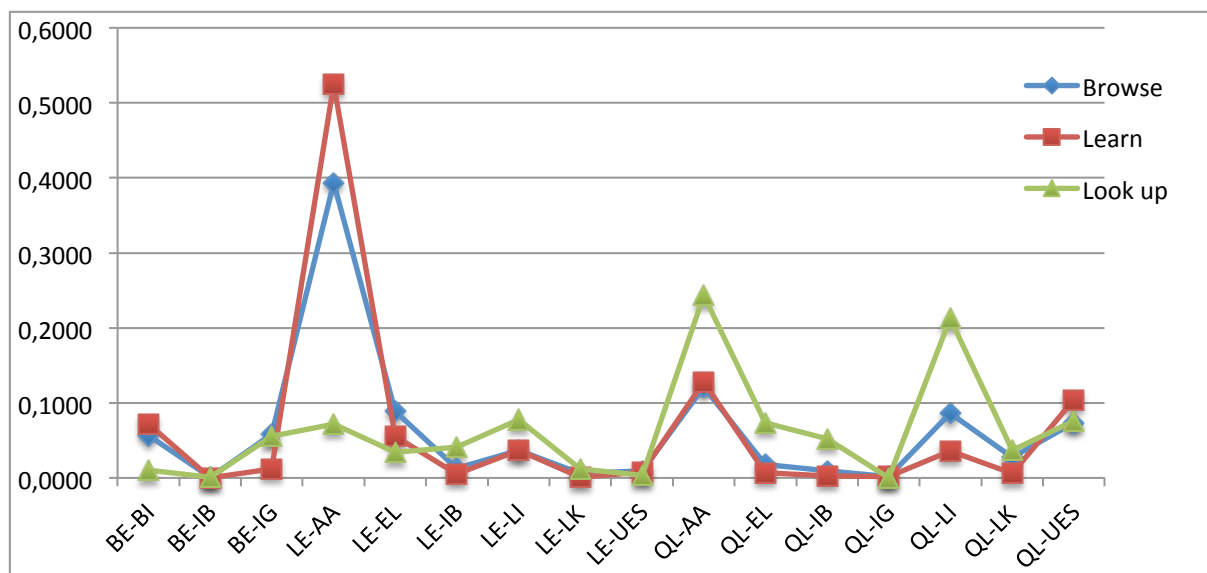


Abbildung 11-17 Mittelwerte der relativen Häufigkeiten der Videolabel-Kombinationen, alle TT, alle VPs

Gerade bei *look up* dominieren auch hier die orientierenden Aktionen. Bei LE-LI, QL-AA, QL-EL, QL-IB und QL-LI liegen die Werte über denen der anderen beiden Aufgabentypen. Alle genannten Labels betreffen nicht so stark der Informationsaufnahme zuzurechnende Nutzeraktionen. Die Kurven von *learn* und *casual leisure* zeigen einen eher ähnlichen Verlauf. Deshalb wurde wieder ein Vergleich der zentralen Tendenz der Verteilungen durchgeführt, diesmal unter Verwendung der relativen Häufigkeiten

<sup>51</sup> Mittelwerte der relativen Häufigkeiten pro Aufgabentyp.

der Videolabelkombinationen respektive der Berücksichtigung der verschiedenen Inhaltstypen.

Eine Analyse der zentralen Tendenz zwischen den drei Stichproben (entsprechen den drei Aufgabentypen) mit Hilfe des Kruskal-Wallis-Test zeigt, dass auch unter Verwendung der an der Anzahl der Interaktionen normierten Nutzeraktionen keine durchgehende Gleichverteilung über die drei Aufgabentypen vorliegt:

Tabelle 11-13 Kruskal-Wallis-Test, Verteilung der  $h_n$  über alle TT

Übersicht über Hypothesentest				
	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von BE-BI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
2	Die Verteilung von BE-IG ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,193	Nullhypothese behalten.
3	Die Verteilung von LE-AA ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
4	Die Verteilung von LE-EL ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,003	Nullhypothese ablehnen.
5	Die Verteilung von LE-LI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,331	Nullhypothese behalten.
6	Die Verteilung von LE-UES ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,092	Nullhypothese behalten.
7	Die Verteilung von QL-AA ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,058	Nullhypothese behalten.
8	Die Verteilung von QL-EL ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
9	Die Verteilung von QL-LI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
10	Die Verteilung von QL-LK ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,270	Nullhypothese behalten.
11	Die Verteilung von QL-UES ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,223	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

Wie oft Bilder betrachtet, Text und Einleitung gelesen und Einleitung und Inhaltsverzeichnis gescannt wurde ist auch bei Verwendung der relativen Häufigkeiten in den Stichproben nicht gleichverteilt.

Anschließend wird deshalb wieder ein Mehrfachvergleich (Games-Howell-Test) durchgeführt, um die einzelnen Unterschiede zwischen den Gruppen genauer zu erfahren. Es werden die fünf durch den Kruskal-Wallis-Test ermittelten unterschiedlich über die drei Stichproben verteilten Variablen getestet. Demnach muss das Signifikanzniveau entsprechend angepasst werden (10%/5). Es gilt also eine maximal erlaubte Irrtumswahrscheinlichkeit von  $p = 0,02$  bzw. 2% für die Ablehnung von  $H_0$ .

Tabelle 11-14 Games-Howell-Test, Mehrfachvergleich,  $h_n$

Mehrfachvergleiche							
Games-Howell							
Abhängige Variable	(I) V1	(J) V1	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
BE-BI	1	2	-,0157	,0181	,668	-,0603	,0289
		3	,0460 <sup>*</sup>	,0138	<b>,010</b>	,0107	,0813
	2	1	,0156	,0182	,668	-,0289	,0603
		3	,0617 <sup>*</sup>	,0129	<b>,000</b>	,0290	,0944
	3	1	-,0460 <sup>*</sup>	,0139	<b>,010</b>	-,0813	-,0107
		2	-,0617 <sup>*</sup>	,0129	<b>,000</b>	-,0944	-,0290
LE-AA	1	2	-,1320	,0699	,160	-,3048	,0407
		3	,3211 <sup>*</sup>	,0612	<b>,000</b>	,1665	,4756
	2	1	,1320	,0699	,160	-,0407	,3048
		3	,4531 <sup>*</sup>	,0458	<b>,000</b>	,3388	,5673
	3	1	-,3211 <sup>*</sup>	,0612	<b>,000</b>	-,4756	-,1665
		2	-,4531 <sup>*</sup>	,0458	<b>,000</b>	-,5673	-,3388
LE-EL	1	2	,0339	,0199	,218	-,0149	,0829
		3	,05497	,0229	,058	-,0015	,1114
	2	1	-,03399	,0199	,218	-,0829	,0149
		3	,0209	,0217	,603	-,0326	,0745
	3	1	-,0549	,0229	,058	-,1115	,0015
		2	-,0209	,0217	,603	-,0745	,0325
QL-EL	1	2	,0116	,0058	,133	-,0029	,0262
		3	-,0554 <sup>*</sup>	,0200	,032	-,1065	-,0045
	2	1	-,0117	,0058	,133	-,0262	,0029
		3	-,0672 <sup>*</sup>	,0194	<b>,008</b>	-,1172	-,0171
	3	1	,0555 <sup>*</sup>	,0200	,032	,0045	,1065
		2	,0672 <sup>*</sup>	,0195	<b>,008</b>	,0171	,1172
QL-LI	1	2	,0502 <sup>*</sup>	,0185	,033	,0036	,0967
		3	-,1266 <sup>*</sup>	,0283	<b>,000</b>	-,1965	-,0565
	2	1	-,0502 <sup>*</sup>	,0184	,033	-,0967	-,0037
		3	-,1767 <sup>*</sup>	,0237	<b>,000</b>	-,2371	-,1164
	3	1	,1265 <sup>*</sup>	,0283	<b>,000</b>	,0565	,1966

	2	,1767*	,0238	,000	,1163	,2371
*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.						

Der Mehrfachvergleich im Rahmen der Varianzanalyse zeigt teilweise statistisch signifikante Unterschiede zwischen den Mittelwerten der Stichproben, vor allem bei den Label-Kombinationen BE-BI (Bilder werden betrachtet), LE-AA (Ein Artikelabsatz wird gelesen) und QL-LI (Querlesen des Inhaltsverzeichnisses) und QL-EL (Querlesen der Einleitung) zwischen *learn* (hier codiert als 2) und *look up* (hier codiert als 3) bzw. zwischen *casual-leisure* (hier codiert als 1) und *look up*. Die Suchstrategien zwischen *learn* und *casual-leisure* unterscheiden sich hier nicht statistisch signifikant.

Interessant ist zusammenfassend, dass sich die Informationsrezeption bei der „Fakten-suche“ (in dieser Arbeit definiert durch das stark *endliche* und *wohldefinierte* Suchziel) deutlich von den beiden anderen Aufgabentypen abgrenzt. Die Suchstrategien bei *learn* und *casual-leisure* unterscheiden sich *beide* gegen *look up* stärker als untereinander. Wie bereits bei den vorangegangenen Untersuchungen zeigt sich der Unterschied vor allem darin, ob verstärkt Information aufgenommen wurde (in Form von Text-Lektüre und der Berücksichtigung von Bildern) und darin, wie viel Zeit der Nutzer damit verbrachte, sich vor allem einen Überblick zu verschaffen (lesen und querlesen des Inhaltsverzeichnisses, der Überschriften und der Einleitung).

### 11.3.6 Analyse der Verteilungen

Um die Verteilung der während der Suche statistisch signifikant unterschiedlich stark berücksichtigten Inhaltselemente genauer zu analysieren wurden grafische Darstellungen hinzugezogen. Das Ziel dieser Analyse ist es, ein detaillierteres Bild zu erhalten, wie die Verwendung der Inhalte verteilt ist.

Einen Überblick gibt zunächst ein gestapeltes Balkendiagramm (Abbildung 11-19). Hier sind die durch den Kruskal-Wallis-Test ermittelten (siehe Tabelle 11-13) zwischen den Aufgabentypen unterschiedlich verteilten Nutzeraktionen (Verwendung der normierten Werte) dargestellt (Mittelwerte der relativen Häufigkeiten). Gut zu erkennen ist hier wiederum der hohe Lektüre-Anteil bei 2 (*learn*) und auch bei 3 (*casual leisure*). Bei *look up* (1) dagegen stellt das scannen des Inhaltsverzeichnisses (QL-LI) den größten Teil dar. Die Einleitung wurde bei *look up* ebenfalls verhältnismäßig viel gelesen und



vor allem gescannt (QL-EL). D.h. Einleitung und Inhaltsverzeichnis wurden von fast jedem Probanden berücksichtigt und mindestens gescannt.

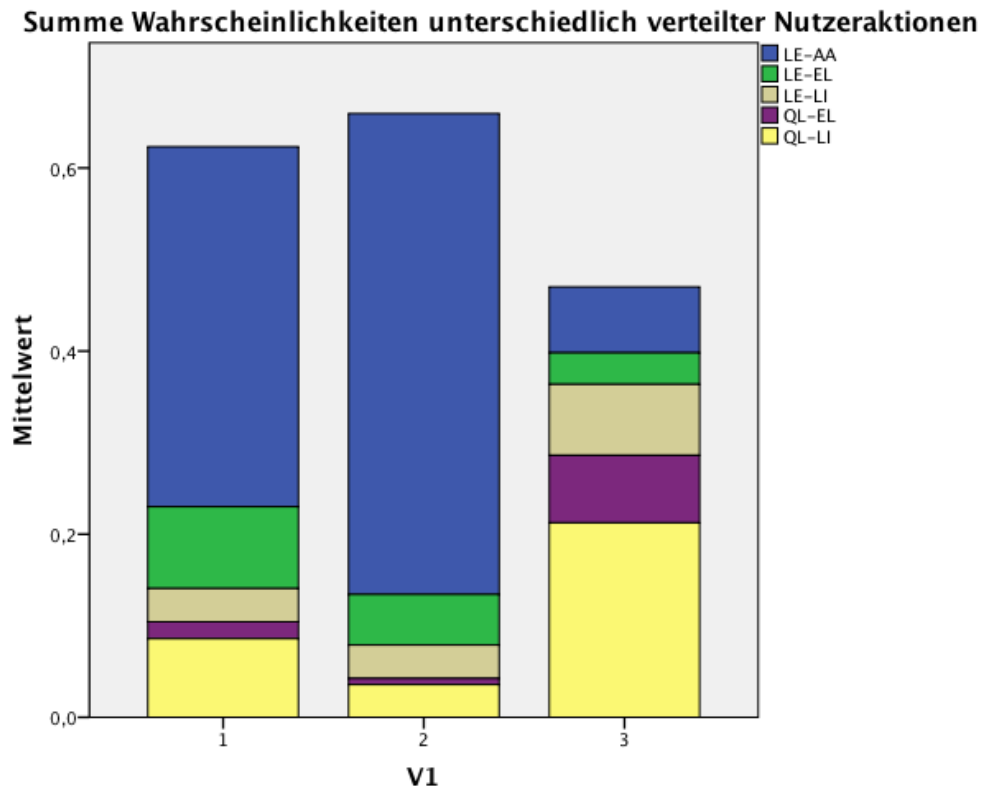


Abbildung 11-18 Balkendiagramm der Wahrscheinlichkeiten/Nutzeraktionen

Die folgenden Histogramme zeigen die drei Videolabelkombinationen, die gerade zwischen *look up* und den beiden anderen Aufgaben (diese mit der Gemeinsamkeit des weniger stark endlichen und definierten Suchziels) am unterschiedlichsten auftreten (BE-BI, LE-AA und QL-EL). Hier sind die relativen Häufigkeiten von  $h_n(\text{Videolabel}) = 0 - 1$  dargestellt.

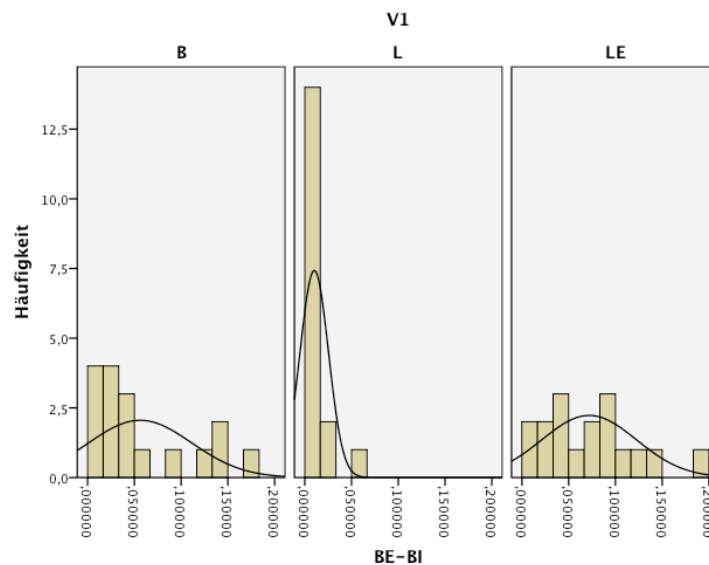


Abbildung 11-19 Histogramm BE-BI Häufigkeit der relativen Häufigkeiten, alle drei TT

Das Betrachten von Bildern ist beim TT *look up* (L) nur mit sehr geringer Wahrscheinlichkeit zu erwarten, bei *casual leisure* (B) und *learn* (LE) dagegen verteilen sich die relativen Häufigkeiten von 0% bis ca. 20 % relativ gleichmäßig, wogegen bei *casual leisure* die Streuung ebenfalls größer als bei *look up* ist, allerdings mit einer linkslastigen Verteilung. Das verdeutlicht, dass die Streuung bei den „freien“ Aufgaben relativ groß ist. Das ist nach den Theorien und Ergebnissen aus Kapitel 1 auch auf die unterschiedliche Eignung und subjektive Verwendung von Bildern zurückzuführen.

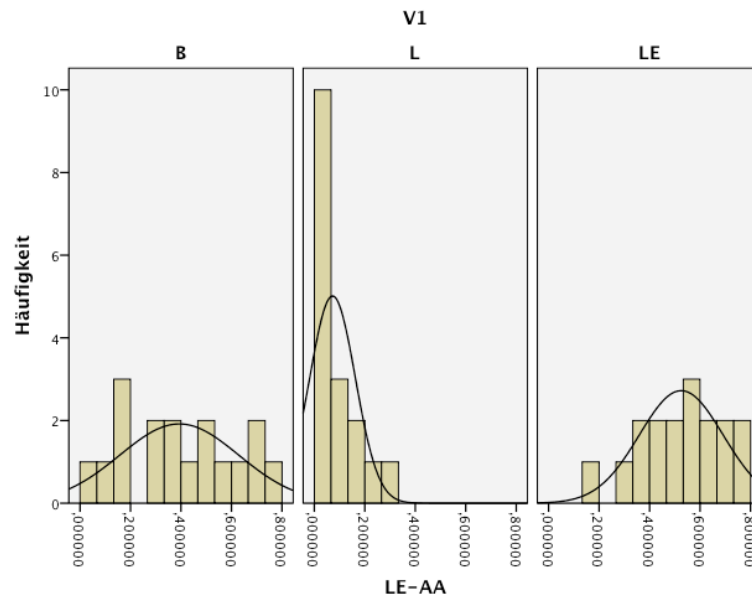


Abbildung 11-20 Histogramm LE-AA Häufigkeit der relativen Häufigkeiten, alle drei TT

Hier (Abbildung 11-20) ist zu erkennen, dass die Verteilung der relativen Häufigkeiten, dass ein Text gelesen wird, bei *look up* vor allem im Bereich zwischen 0% und 20% liegt, wogegen bei den *learn*-Aufgaben eine rechtslastige Verteilung, also hohe Wahrscheinlichkeiten für das Lesen von Text zu erkennen sind. Bei *casual-leisure* ist die Streuung am größten; die relativen Häufigkeiten verteilen sich relativ gleichmäßig zwischen  $h_n$  (LE-AA) = 0 und  $h_n$  (LE-AA) = 0,8. Diese im Vergleich zu den beiden anderen Aufgabentypen eher große Streuung ist dadurch zu erklären, dass bei diesem Aufgabentyp der Nutzer das Thema frei wählen konnte und kein Lernergebnis erwartet wurde. Hier wurde deshalb vermutlich vor allem je nach Interesse entweder gelesen oder sich nur überblicksartig informiert. Gefordertes Ziel dieser Aufgabe war es lediglich, einen Zeitraum von fünf Minuten zu überbrücken.

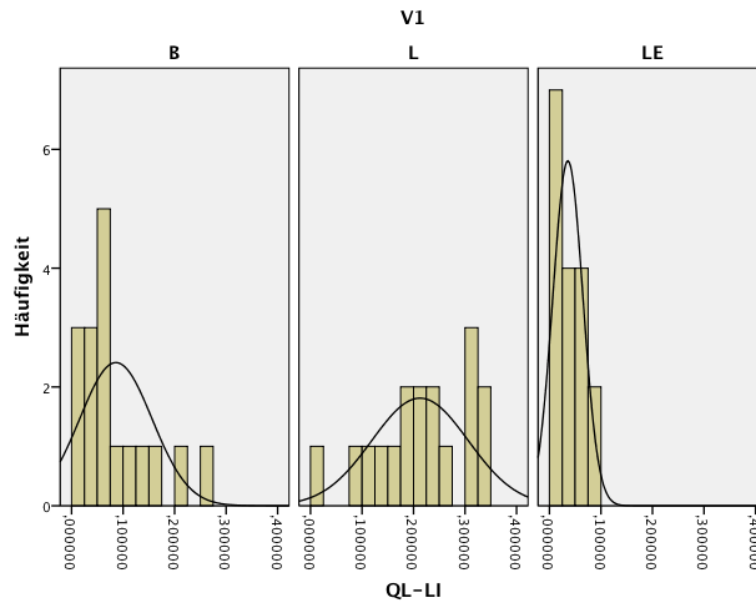


Abbildung 11-21 Histogramm QL-LI Häufigkeit der relativen Häufigkeiten, alle drei TT

Auch beim Scannen einer Liste - hier vor allem dem Inhaltsverzeichnis der Artikel - zeigen die Histogramme deutliche Unterschiede in der Verteilung. Bei den *learn*-Aufgaben liegen die relativen Häufigkeiten alle bei  $\leq 10\%$ . Die Leser haben hier also wenig Zeit für die Orientierung und Übersicht über den Artikel aufgewendet. Bei *look up* dagegen liegt der größere Teil der relativen Häufigkeiten bei über 20%. Hier wurde die Liste/Inhaltsverzeichnis stärker herangezogen. Bei den *casual-leisure*-Aufgaben ist zwar ebenfalls eine linkslastige Verteilung zu erkennen, allerdings nicht so stark wie bei *learn*. Tendenziell wurde auch eher wenig Zeit zur Orientierung verwendet, allerdings sichtbar mehr als bei den *learn* Aufgaben.

Die Histogramme veranschaulichen insgesamt die unterschiedliche Verteilung der relativen Häufigkeiten der Videolabelkombinationen und damit der Informationsrezeption bei den verschiedenen Nutzungsszenarien (hier: Aufgabentypen).

In der Zusammenfassung der verschiedenen Tests und der Berücksichtigung der graphischen Darstellungen zeigt sich als Gesamtergebnis der Auswertung, dass sich bei den *learn* und der *casual-leisure* Szenarien das Verhalten der Nutzer und deren Präfe-

renzen für bestimmte Inhaltselemente in weniger Punkten statistisch signifikant unterscheiden als beide Typen gegenüber dem Fakten-Retrieval.

Die wichtigste Erkenntnis ist dabei, dass gerade beim Lesen von Texten (LE-AA) und bei der Berücksichtigung von Bildern (BE-BI) bei allen Tests zwischen *look up* und den anderen beiden Szenarien (*learn* und *casual-leisure*) statistisch signifikante Unterschiede nachzuweisen sind.

Mehr als einmal auftretende statistisch signifikante Unterschiede (bei den Mittelwerten so wie dem Kruskal-Wallis-Test) liegen bei den Elementen Einleitung und Liste (Inhaltsverzeichnis) vor.

Tabelle 11-15 Zusammenfassung der Test-Ergebnisse. Spalte 1-3: Ergebnisse des t-Tests, unter Verwendung der absoluten ( $n$ ) und der relativen Häufigkeiten ( $h_n$ ). Spalte 5&6: Ergebnisse des Kruskal-Wallis-Tests, unter Verwendung der relativen Häufigkeiten ( $h_n$ )

	Games Howell B vs. LE		Games Howell B vs. L		Games Howell L vs. LE		K-W Test	
	$n$	$h_n$	$n$	$h_n$	$n$	$h_n$	B, LE, L ( $n$ )	B, LE, L ( $h_n$ )
BE-BI	x	x	0,000	0,010	0,00	0,000	0,000	0,000
BE-IB	x	x	x	x	x	x	x	x
BE-IG	x	x	x	x	x	x	x	x
LE-AA	x	x	0,000	0,000	0,000	0,000	0,000	0,000
LE-EL	x	x	0,000	x	0,001	x	0,000	0,003
LE-IB	x	x	x	x	x	x	x	x
LE-LI	x	x	x	x	x	x	x	x
LE-LK	x	x	x	x	x	x	0,015	x
LE-UES	x	x	0,011	x	x	x	0,016	x
QL-AA	x	x	x	x	x	x	x	x
QL-EL	x	x	x	x	x	x	x	0,000
QL-IB	x	x	x	x	x	x	0,003	x
QL-IG	x	x	x	x	x	x	x	x
QL-LI	x	x	x	0,000	x	0,000	0,022	0,000
QL-LK	x	x	x	x	x	x	x	x
QL-UES	x	x	x	x	0,001	x	0,000	x

Die Zusammenfassung in Tabelle 11-15 zeigt nur die entsprechend der jeweils festgelegten maximal erlaubten Irrtumswahrscheinlichkeit statistisch signifikanten Ergebnisse. Demnach sind keine unterschiedlichen Verteilungen bzw. statistisch signifikante Mittelwertdifferenzen zwischen *learn* und *casual leisure* festzustellen, die sicher genug sind, um auf darauf schließen zu können, dass diese Unterschiede nicht zufällig sind.

Zwischen den beiden Suchtypen mit weniger endlichem und weniger stark definierten Suchziel (*learn* und *casual leisure*, zur genaueren Definition siehe auch 11.2.2) und dem Faktenretrieval (*look up*) existieren so starke Unterschiede, dass diese mit einer relativ hohen Sicherheit auch auf die Grundgesamtheit zutreffen. Nachfolgende Abbildung 11-22 zeigt noch einmal die unterschiedliche Streuung der drei Nutzeraktionen Lektüre von Text, Berücksichtigung von Bildern und Lektüre der Einleitung in Abhängigkeit des Aufgabentyps.

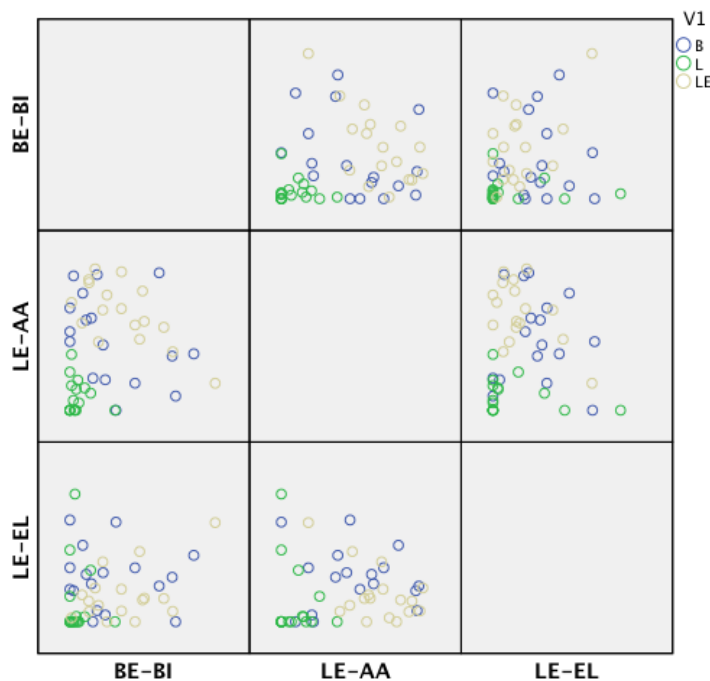


Abbildung 11-22 Matrix der drei am unterschiedlichsten verteilten Nutzeraktionen, relative Häufigkeiten

Diese Ergebnisse bestätigen die Hypothese, dass sich die Nutzerpräferenzen bzw. die subjektive situative Relevanz von Inhaltselementen bei der Suche in der Wikipedia abhängig vom Suchziel und der Motivation unterscheiden. Die Unterschiede konzentrieren sich vor allem auf die Verwendung von überblicksartigen Elementen vs. ausführlicher Informationsaufnahme (hier vor allem der Artikel-Text und Bilder).

Da die in dieser Arbeit durchgeführten Untersuchungen als mögliche Grundlage für ein adaptives Information Retrieval zur Auswahl der geeigneten Inhaltselemente in Abhängigkeit des Aufgabentyps dienen, wird im nächsten Schritt untersucht, ob sich

die Datensätze der Tasks, die die Suchstrategien und Präferenzen der Nutzer repräsentieren, so stark unterscheiden, dass eine automatische Klassifikation möglich ist.

### 11.3.7 Automatische Klassifikation der Daten - taskweise

Es wurde untersucht, ob es möglich ist, die Daten (= taskweise Angaben zu den Wahrscheinlichkeiten bzw. Häufigkeiten der Labelkombinationen) automatisch zu klassifizieren. Verschiedene Vorgehensweisen zur automatischen Klassifikation der Daten kamen zum Einsatz unter Benutzung der freien Software *Weka* (*Waikato Environment for Knowledge Analysis*), die verschiedene tools zur Daten-Analyse und *predictive modeling* bereitstellt und *IBM SPSS Statistics 21*. Es wurden verschiedene Modelle getestet.

Verwendet man den *Naive Bayes Klassifikator*<sup>52</sup> (Weka) und lässt diesen die Tasks zunächst bei Verwendung *aller* Labelkombinationen und deren *absoluten* Häufigkeiten klassifizieren, so erreicht der Klassifikator gute 90% *correctly classified instances*.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.7	0	1	0.7	0.824	0.915	B
1	0	1	1	1	1	L
1	0.143	0.76	1	0.864	0.936	LE
Weighted Avg.						
	0.902	0.044	0.925	0.902	0.9	0.952

Es ist allerdings so, dass bei den absoluten Häufigkeiten in den Daten auch deshalb Unterschiede bestehen, da die Tasks unterschiedlich lang sind. *Learn*- und *browse*-Tasks haben teilweise ähnliche Anzahlen von Frames, da für die *learn*-Tasks theoretisch die selbe Zeit zur Verfügung stand wie für *casual-leisure*, allerdings der Task von der VP beendet werden konnte, wenn das Informationsbedürfnis nach subjektiver Einschätzung ausreichend erfüllt war.

Wie die *Confusion Matrix* zeigt, machte der Klassifikator auch „nur“ bei der Zuordnung von *casual-leisure*-Tasks einen Fehler und klassifizierte diese sechsmal als *learn*-Task.

<sup>52</sup> Dieser geht von der Unabhängigkeit der einzelnen Parameter aus. Mit Hilfe einer Menge an Attributen (dem Auftreten bestimmter Videolabel-Kombinationen)  $x_1 \dots x_n$  soll ein Attribut  $V$  (in diesem Fall der Tasktyp) vorhergesagt werden. Dabei werden für einen bestimmter Attribut-Vektor  $E = [x_1, \dots, x_n]$  alle Wahrscheinlichkeiten berechnet und die Größte ausgewählt [siehe auch Klawonn, 2004].

=== Confusion Matrix ===

a b c <-- classified as

14 0 6 | a = B

0 22 0 | b = L

0 0 19 | c = LE

Die Prozentsätze der richtig erkannten Tasks, wenn anhand nur *einer* (der nach dem Kruskal-Wallis-Test *nicht* gleichverteilten) Label-Kombination mit dem *Naive Bayes Klassifikator* klassifiziert wurde, liegt zwischen 27,86 % und 73,77 %. Es wurden ebenfalls die absoluten Häufigkeiten verwendet. Interessant ist, dass die Tasks anhand der Verwendung von Bildern am besten klassifiziert werden konnten, was auch die situative Relevanz von Bildern bestärkt (siehe dazu auch Kapitel 1).

Tabelle 11-16 Ergebnisse der automatischen Klassifikation unter Verwendung nur einer Videolabel-Kombination

BE-BI	Correctly Classified Instances	45	<b>73.7705 %</b>
	Incorrectly Classified Instances	16	26.2295 %
LE-AA	Correctly Classified Instances	42	68.8525 %
	Incorrectly Classified Instances	19	31.1475 %
LE-EL	Correctly Classified Instances	26	42.623 %
	Incorrectly Classified Instances	35	57.377 %
QL-EL	Correctly Classified Instances	17	27.8689 %
	Incorrectly Classified Instances	44	72.1311 %
QL-LI	Correctly Classified Instances	29	47.541 %
	Incorrectly Classified Instances	32	52.459 %

Da die Verwendung absoluter Häufigkeiten dem Klassifikator auf Grund der unterschiedlichen Längen bzw. der verschiedenen Anzahlen der 50-ms-Frames die Zuordnung insofern verfremdet, als dass die Längen der Tasks nicht als *feature* zur Klassifikation herangezogen werden sollte<sup>53</sup>, ist es sinnvoller, auch bei der automatischen Klassifikation der Daten mit den relativen Häufigkeiten zu arbeiten.

<sup>53</sup> Die absolute Dauer wurde in dieser Arbeit nicht als Einflussfaktor berücksichtigt. In den Experimenten wurden Zeitbegrenzungen vorgegeben, um den Nutzer zu einer selektiven Vorgehensweise zu animieren.



Außerdem wurden weiterhin dem Klassifikator nur die durch den Kruskal-Wallis-Test (siehe Tabelle 11-10) als unterschiedlich verteilten Videolabel-Kombinationen bereitgestellt.

Am besten schnitt dabei die „*Simple Logistic Regression*“ ab mit einer Rate von 76,47% richtig klassifizierten Datensätzen.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.588	0.147	0.667	0.588	0.625	0.801	B
0.765	0.147	0.722	0.765	0.743	0.888	LE
0.941	0.059	0.889	0.941	0.914	0.984	L

Weighted Avg. 0.765 0.118 0.759 0.765 0.761 0.891

Interessant ist hierbei, dass wiederum die *casual-leisure*-Aufgaben (B) am schlechtesten zugeordnet werden konnte. Das zeigt auch die *Confusion Matrix*, B wurde insgesamt siebenmal falsch zugeordnet, L dagegen nur einmal.

=== Confusion Matrix ===

a b c <-- classified as

10 5 2 | a = B

4 13 0 | b = LE

1 0 16 | c = L

Mit 72,55 % richtig zugeordneten Datensätzen schnitt auch das Modell „*attributeSelection.BestFirst -D 1 -N 5*“ relativ gut ab. Bei diesem Modell wird nur eine ausgewählte Gruppe von Features zur Klassifikation verwendet.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.647	0.206	0.611	0.647	0.629	0.711	B
0.824	0.176	0.7	0.824	0.757	0.814	LE
0.706	0.029	0.923	0.706	0.8	0.928	L
Weighted Avg.						
	0.725	0.137	0.745	0.725	0.728	0.818

Wieder fiel die Zuordnung der *casual-leisure*-Aufgaben am schlechtesten aus (sechs falsche Zuordnungen), diesmal knapp gefolgt von *look up* (fünf mal falsch zugeordnet).

=== Confusion Matrix ===

```

a b c <-- classified as
11 5 1 | a = B
3 14 0 | b = LE
4 1 12 | c = L

```

Als dritter Klassifikator wurde ein Entscheidungsbaum verwendet. Dieser sortierte die Datensätze ebenso gut wie der Klassifikator mit der *attribute selection*, und zwar zu 72,55 % richtig.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.647	0.206	0.611	0.647	0.629	0.728	B
0.647	0.118	0.733	0.647	0.688	0.839	LE
0.882	0.088	0.833	0.882	0.857	0.959	L
Weighted Avg.						
	0.725	0.137	0.726	0.725	0.724	0.842

Hier wurden die *casual-leisure* und *learn* Aufgaben jeweils sechsmal falsch zugeordnet, bei *look up* gab es nur zwei falsche Sortierungen.

=== Confusion Matrix ===

```
a b c <-- classified as
11 3 3 | a = B
6 11 0 | b = LE
1 1 15 | c = L
```

Insgesamt zeigen die Ergebnisse, dass das Verhalten der Nutzer bei der Suche, das durch die Datensätze repräsentiert wird, relativ gut klassifiziert werden kann. Die Ergebnisse zw. 70% und 80% noch nicht optimal, aber eine gute Grundlage.

In den Phasen einer Suche ändert sich die Verteilung der Videolabels bzw. der Nutzeraktionen möglicherweise, berücksichtigt man die Theorie zu den Suchphasen von [Kuhlthau, 1991].

Da die Aktionen *nicht* gleichverteilt sind am Anfang eines Tasks (entspricht hier den ersten 50 Frames à 50 ms), ist zu vermuten, dass in den *learn* und *browse* Aufgaben schneller mit der Informationsaufnahme begonnen wird als bei den *look-up* tasks (siehe Kruskal-Wallis-Test Anhang 5.8).

Das zeigt auch das Diagramm der Summe<sup>54</sup> der absoluten Häufigkeiten aller VPs der Videolabel-Kombinationen in den ersten 50 Frames.

---

<sup>54</sup> Hier ist keine Normierung der Aktionen notwendig, da für alle die gleiche Anzahl der Nutzerinteraktionen (n = 50) gegeben ist.

n (frames)

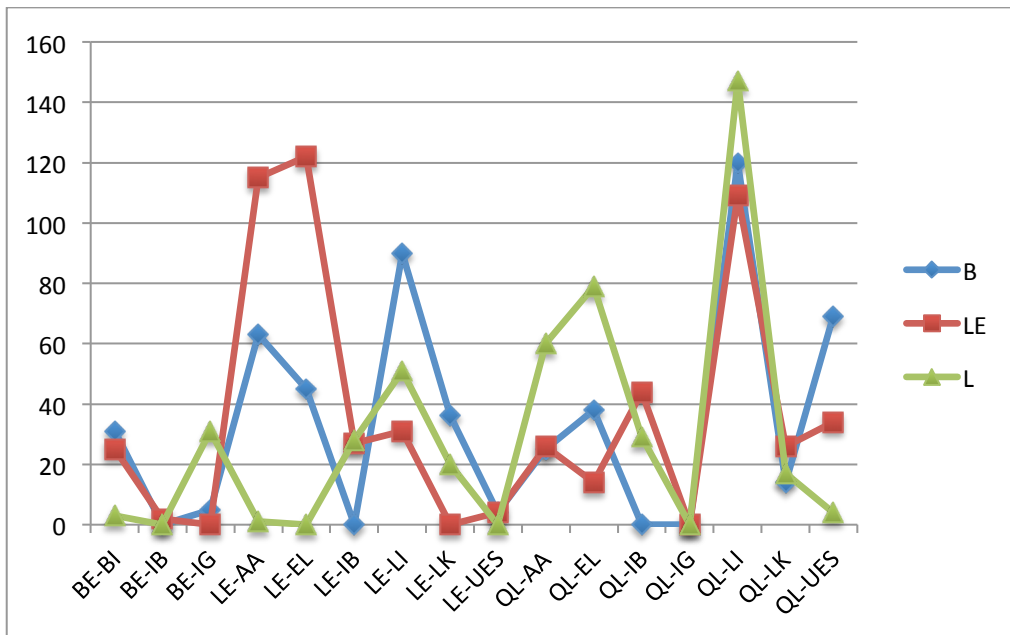


Abbildung 11-23 Summe der absoluten Häufigkeiten der Videolabel-Kombinationen (50ms frames) aller VPs

Die Grafik zeigt über die ersten 50 Frames, dass bei den *look up* Aufgaben die Einleitung zum Einstieg nicht gelesen wurde, auch keine Artikelabsätze. Die Liste, in diesem Fall das Inhaltsverzeichnis, wurde in allen drei Fällen stark gescannt. Bei den *learn* Aufgaben wurde bereits in den ersten 50 x 50 ms sowohl die Einleitung als auch der Artikeltext gelesen. Der Kruskal-Wallis-Test zeigt, dass QL-UES, QL-EL und LE-AA in den ersten 50 Frames über die drei Aufgabentypen nicht gleichverteilt sind (siehe Anhang 5.8). Diese Werte zeigen, dass sich die Suchstrategien der Nutzer abhängig von der Aufgabe bereits am Anfang eines Tasks unterscheiden.

### 11.3.8 Weitere Einflussfaktoren

Es wurden entsprechend der Literatur und dem in dieser Arbeit erstellten Modell weitere Faktoren auf ihren Einfluss auf das Suchverhalten untersucht.

Das Geschlecht der Probanden spielte in dieser Studie keine Rolle, eine Untersuchung der Mittelwerte zwischen der Gruppe der männlichen und der weiblichen Versuchspersonen mittels des t-Tests<sup>55</sup> ergab nur für das Querlesen von Überschriften einen statistisch signifikanten Unterschied (siehe Anhang 5.9).

<sup>55</sup> Anwendung des t-Tests trotz nicht normalverteilter Daten, da Gruppengröße  $n > 30$  [Leonhart, 2013].

Die Auswertung der Angaben zum emotionalen Zustand des Nutzers mittels des *Self Assessment Manikin* [Bradley und Lang, 1994] zeigte zunächst, dass sich der emotionale Zustand der Testpersonen während des *gesamten* Tests wenig änderte. Die Auswertung wurde weiterhin *taskabhängig* vorgenommen.

Für die Auswertung der Bilder des SAM wurden entsprechend der Beschreibung von [Bradley und Lang, 1994] bezüglich der bipolaren Wortpaare zur qualitativen Beschreibung von Information (siehe Tabelle 11-17) Werte vergeben: für Bild 1 in der Dimension *valence* fünf Punkte absteigend bis Bild 5, das mit einem Punkt bewertet wurde. Bei der zweiten Dimension *arousal* erfolgte die Punktevergabe umgekehrt: Bild 1 wurde mit einem Punkt bewertet, aufsteigend bis Bild 5, für das fünf Punkte vergeben wurden. Hintergrund dabei ist, dass die erste Dimension Gefühle wie „happy“ oder „pleased“ widerspiegeln – also je intensiver die erste Dimension ausgeprägt ist, desto positiver fühlt der Proband sich (siehe dazu Faktorenladung Tabelle 11-17).

Tabelle 11-17 [Bradley und Lang, 1994, S.50] Faktorenladung der Wortpaare des Semantischen Differential nach Reduktion auf die drei Bereiche "Valence", "Arousal" und "Dominance"

	Factor 1 „Pleasure“	Factor 2 „Arousal“	Factor 3 „Dominance“
Unhappy-Happy	0,914	0,063	0,148
Annoyed-Pleased	0,883	0,068	0,158
Unsatisfied-Satisfied	0,868	0,144	0,114
Melancholic-Content	0,725	0,095	0,056
Despairing-Hopeful	0,858	0,063	0,078
Bored-Relaxed	0,580	0,372	0,234
Relaxed-Stimulated	-0,211	0,774	0,052
Calm-Excited	-0,181	0,793	0,056
Sluggish-Frenzied	0,268	0,771	0,005
Dull-Jittery	-0,211	0,793	0,121
Sleepy-Wide awake	-0,046	0,810	0,047
Unaroused-Aroused	0,051	0,827	0,127
Controlled-Controlling	0,262	0,192	-0,673
Influenced-Influential	0,292	0,089	-0,618
Carred for-In control	-0,090	0,198	-0,626
Awed-Important	0,199	-0,040	-0,301
Submissive-Dominant	0,195	0,306	-0,695
Guided-Autonomous	0,161	0,100	-0,479
Amount of variance accounted for:	24,6	23,12	12,18

Die zweite Dimension *arousal*, am besten mit „Anspannung“ zu übersetzen, versucht den Grad an Anspannung einzufangen, dem der Proband in diesem Moment ausgesetzt ist. In diesem Fall wird die Dimension umso niedriger bewertet, je stärker sie ausgeprägt ist (siehe dazu Faktorenladung Tabelle 11-17).

Personen, die in der Dimension *valence* nur zwei oder einen Punkt angegeben haben, werden entsprechend der bipolaren Wortpaare in Tabelle 11-17 in die Kategorie „negative Stimmung“ eingeteilt, die anderen in „positive Stimmung“, gleiches gilt für die Dimension *arousal*.

Bei nur sieben Aufgaben änderten sich die Angaben in der Dimension *arousal* um mehr als einen Punkt, bei immerhin elf Tasks gingen die Angaben bei *pleasure* um mehr als einen Punkt hinauf oder hinunter. Nur zweimal verbesserte die *casual leisure* Aufgabe die Stimmung des Probanden in der Kategorie *pleasure*. Demnach ist es in dieser Laborsituation offensichtlich schwer, affektive und emotionale Reaktionen zu erzeugen und einzufangen. Ausschlaggebend für die Erfassung der emotionalen Lage war initiativ vor allem die Annahme, dass die VPs zu Testbeginn unterschiedlicher Stimmung sind, bzw. auch die Testsituation im Labor dazu beitragen kann, bestimmte emotionale Reaktionen hervorzurufen (gestresst, angespannt, durch Beschäftigung mit persönlich interessanten Themen in den *casual-leisure* Aufgaben aber auch entspannt und fröhlich ...). Diese Reaktionen blieben allerdings in einem schwachen Bereich.

Insgesamt zeigte sich, dass beide Verteilungen in den Dimensionen *arousal* und *valence* die häufigsten Angaben bei der jeweils zweipositivsten Ausprägung hatten. In beiden Dimensionen gab niemand die letzte Kategorie an.

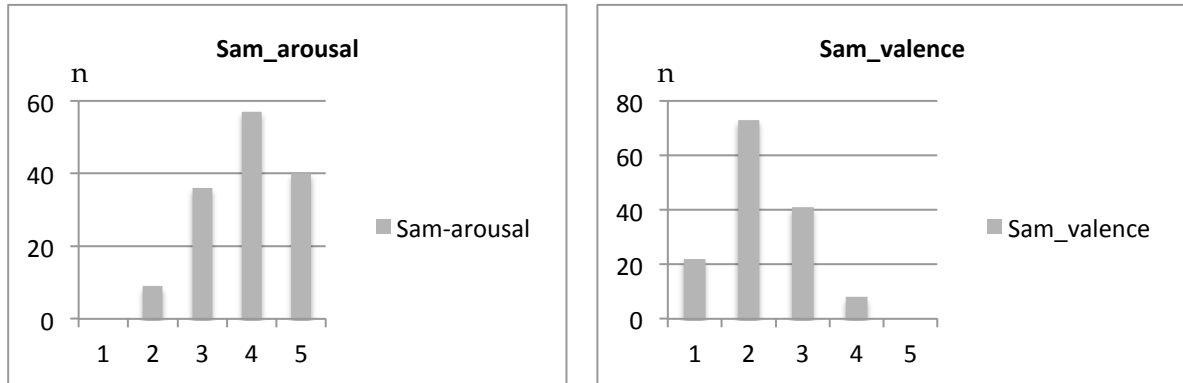


Abbildung 11-24 emotionale Selbsteinschätzung der Probanden anhand des im Fragebogen verwendeten SAM. (arousal: 1 = erregt, 5 = ruhig, valenz: 1 = angenehm, 5 = unangenehm)

Die dritte Dimension des SAM (*dominance*) wurde zu den Auswertungen nicht weiter herangezogen, da sich herausstellte, dass einem großen Teil der Testpersonen die Bedeutung dieser Dimension nicht klar war. Da für den Testablauf vereinbart wurde, dass keine weitere Kommunikation mit dem Probanden stattfinden sollte, konnte dies auch nicht durch Nachfragen geklärt werden und es besteht der Verdacht, dass die erfassten Daten der Dimension *dominance* nicht durchgehend valide sind.

Die Überprüfung der Korrelation der Werte von *pleasure* and *arousal* mit den Häufigkeiten der jeweils betrachteten Inhaltselemente taskweise ergab keine Ergebnisse, die statistisch signifikant waren (Korrelationskoeffizient zwischen -0,2 und + 0,2). Es wurde jeweils die Korrelation nach Pearson überprüft.

Die Zufriedenheit der Nutzer korreliert positiv mit dem subjektiven Erfolg der Suche, der in den Fragebögen angegeben wird (siehe Tabelle 11-18). Das bedeutet, je erfolgreicher die Testpersonen ihre Suche bewerteten, desto zufriedener waren sie auch mit dem Suchverlauf. Ein erfolgreicher Suchverlauf ist dementsprechend wichtig für die Nutzerzufriedenheit (nicht nur inhaltlicher Erfolg). Der hohe Signifikanzwert von 0,01 besagt auch eine große Wahrscheinlichkeit, dass das auf die Grundgesamtheit zutrifft.

Tabelle 11-18 Korrelation nach Pearson Erfolg und Zufriedenheit mit der Suche

Korrelationen		Erfolg	Zufrieden_Verlauf
Erfolg	Korrelation nach Pearson	1	,659**
	Signifikanz (2-seitig)		,000
	N	102	102
Zufrieden_Verlauf	Korrelation nach Pearson	,659**	1
	Signifikanz (2-seitig)	,000	
	N	102	102

\*\* . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Die Hypothese, dass die Verwendung bestimmter Inhaltselemente auch von der emotionalen Situation der Person abhängen, konnte in diesem Experiment nicht direkt bestätigt werden.

Eine Erklärung dafür ist, dass – wie die Korrelation der emotionalen Lage mit der Zufriedenheit mit den Suchergebnissen zeigt – es sich bei den Angaben, die die VPs mittels SAM machten, um kurze Momentaufnahmen handelt, bei denen keine starken Schwankungen auftraten. In der Konsequenz muss sich eine Untersuchung auf eine grundlegendere emotionale Verfassung über einen längeren Zeitraum konzentrieren. Die Suche in der Wikipedia in einer Laborstudie hatte keinen Einfluss auf den Nutzer, der stark genug war, um tatsächliche affektive Reaktionen hervorzurufen.

Ähnliches bemerken auch Nahl und Bilal [Nahl und Bilal, 2007] im Zusammenhang damit, dass die Stimmung bzw. der emotionale Zustand einer Person von potentiell sehr vielen Faktoren beeinflusst wird, die teilweise nichts mit der Suche selbst zu tun haben. Eine Stimmung aber dauert in der Regel über einen längeren Zeitraum an. Sie könnte während der Suche möglicherweise durch starke Einflüsse verändert werden, was aber bei den vorliegenden Experimenten nicht der Fall war (nur schwache Änderungen bei der Selbsteinschätzung der emotionalen Stimmung). Wie die Auswertung des SAMs zeigte, war die grundsätzliche Verteilung der emotionalen Verfassung zu Beginn der Studie ohne Extreme.

Die logische Folgerung aus den Erfahrungen und Daten dieser Arbeit sowie der Theorie ist, dass der emotionale Einfluss bei der Suche in der Wikipedia eher bei einem



Langzeitphänomen wie der Stimmung der Nutzer zu suchen ist, es sei denn, es werden subjektiv emotional aufgeladene Inhalte (entsprechend der kognitiven Emotionstheorie [Scherer, 1990]) verwendet, was hier nicht der Fall war.

Im Gegensatz zu Untersuchungen bei Computerspielen [siehe Picard, 1998], wo Stress und Ärger schnell und stark entstehen können, ist das Gebiet der Informationssuche in der Wikipedia bei den hier verwendeten, allgemein eher *neutralen* Themen nicht so stark von Emotionen geprägt. Ärgernisse über Inhalte oder Navigationsprobleme, die Grund für affektive Reaktionen sein könnten, wurden nicht berichtet. Der Spaßfaktor bei der Suche wurde im Mittel mit 2,73 (bei max. fünf möglichen Punkten) angegeben. Es besteht nur eine äußerst geringe und damit zu vernachlässigende Korrelation zwischen dem angegebenen Spassfaktor einer Aufgabe und den Angaben zur emotionalen Selbstauskunft mittels SAM (Korrelationskoeffizient nach Pearson  $\leq 0,2$ ).

Um Emotionen als Einfluss- und Zielgröße bei der Suche in der Wikipedia zu evaluieren, ist eine über einen längeren Zeitraum und außerhalb des Labors angelegte Studie deshalb zu bevorzugen.

Das Interesse am jeweiligen Thema zeigte *keinen* Einfluss auf die Verwendung der Inhaltselemente, anders allerdings verhält sich das bei der thematischen Vorkenntnis der Probanden. Die Probanden wurden zur Überprüfung in zwei Gruppen eingeteilt: Mit Vorkenntnis bedeutet im *Pre-Task* Fragebogen eine Angabe von über zwei Punkten (Mittelwert aller Probanden ist 2,19 von max. fünf Punkten), keine oder wenig Vorkenntnis gleich oder unter zwei Punkten. Es wurden hier nur zwei Gruppen gebildet, da die Angaben zur Vorkenntnis eine relativ geringe Streuung zeigen (siehe Tabelle 11-19).

Zwischen der Gruppe mit Vorkenntnis ( $n = 23$ ) bestand ein statistisch signifikanter Unterschied (siehe dazu Anhang 5.10) bei der Berücksichtigung des Inhaltsverzeichnisses, sowohl bei der Lektüre als auch beim Querlesen gegenüber der Gruppe ohne oder mit wenig Vorkenntnis ( $n = 36$ ). Dieses Ergebnis ist zwar nicht überraschend, allerdings ein guter Hinweis darauf, dass die jeweilige thematische Versiertheit eines Nutzers unter anderem an der Zeit, die er zur Lektüre des Inhaltsverzeichnisses aufwendet, abgeschätzt werden kann.

Tabelle 11-19 Deskriptive Statistik zur Angabe "thematische Vorkenntnis"

Vorkenntnis		Statistiken
N	Gültig	59
	Fehlend	31
Mittelwert		2,19
Standardabweichung		1,074
Varianz		1,154
Minimum		1
Maximum		5

Bei dieser Auswertung wurden die *casual-leisure* Aufgaben nicht berücksichtigt, da erstens die Themen innerhalb einer Aufgabe wechseln konnten und zweitens bei Erstellung der Angaben im *Pre-Task* Fragebogen das Thema noch nicht bekannt war, demnach also auch keine Angaben zur thematischen Vorkenntnis gemacht werden konnten. Die Vorkenntnis zu einem Thema korreliert mittelstark (Korrelationskoeffizient nach Pearson 0,598, auf dem Niveau von 0,01 signifikant) mit dem Interesse des Probandes (siehe dazu Anhang 5.11).

### 11.3.9 Regressionsanalyse

Es wurde weiter eine binär logistische Regression durchgeführt, um so zu untersuchen, welche einzelnen Nutzeraktionen *Prädiktoren* für die Aufgabentypen sind im Sinne einer Vorhersage der Aufgabenart. Der Vorteil bei der *logistischen* Regression ist, dass nicht davon ausgegangen wird, dass die Veränderungen der unabhängigen Variable *immer* proportional zur abhängigen Variable sind [Leonhart, 2013], denn es ist hier nicht anzunehmen, dass jede einzelnen Zunahme einer Nutzeraktion den Aufgabentyp „ändert“.

Bei dem hier angewendeten Verfahren werden die unabhängigen Variablen (Prädiktoren) schrittweise in das Modell aufgenommen und auf ihre Signifikanz geprüft. Es wurden alle Nutzeraktionen überprüft (eine Vorauswahl wäre anhand der bisherigen Erkenntnisse auch möglich gewesen, allerdings durch die unterschiedlichen Unterscheidungskriterien zwischen den Aufgabentypen möglicherweise zu verlustbehaftet). Der Algorithmus endet, wenn entweder alle Variablen Teil des Modells sind oder auf Grund statistischer Werte keine neuen Variablen mehr aufgenommen werden. Der cut-off Wert für die Variablen wurde auf 0,5 festgelegt.

Dieses Verfahren wurde nur für die Aufgabentypen (und nicht für weitere ursächliche Faktoren) durchgeführt, da sich in den bisherigen Ergebnissen der Eindruck manifestierte, dass hier die Zusammenhänge zu schwach sind um zuverlässige Prädiktoren zu finden (siehe dazu in den einzelnen Kapiteln).

Es wurden jeweils alle Labels auf ihre Eigenschaft als Prädiktor für jeweils einen Aufgabentyp getestet. Die vorhergesagte Variable ist dabei eine *dummy Variable*, die als 1 oder 0 jeweils dafür steht, ob es sich um den gesuchten Aufgabentyp handelt oder nicht.

Es zeigte sich, dass unterschiedliche Nutzeraktionen relevante Prädiktoren für die Aufgabentypen sind. Bei *learn* konnten nach sechs Anpassungsschritten 87,8 % der Aufgaben richtig vorhergesagt werden, bei *casual-leisure* 78,9 und bei *look up* sogar 97,8 %. Dabei waren jeweils verschiedene Prädiktoren relevant (siehe dazu Tabelle 11-20 bis Tabelle 11-22)

Tabelle 11-20 binär logistische Regression *learn*

Klassifizierungstabelle <sup>a</sup>				
	Beobachtet	Vorhergesagt		
		is learn		Prozentsatz der Richtigen
		0	1	
Schritt 1	0	52	8	86,7
	is learn 1	15	15	50,0
	Gesamtprozentsatz			74,4
Schritt 2	0	50	10	83,3
	is learn 1	13	17	56,7
	Gesamtprozentsatz			74,4
Schritt 3	0	53	7	88,3
	is learn 1	8	22	73,3
	Gesamtprozentsatz			83,3
Schritt 4	0	52	8	86,7
	is learn 1	8	22	73,3
	Gesamtprozentsatz			82,2
Schritt 5	0	51	9	85,0
	is learn 1	5	25	83,3
	Gesamtprozentsatz			84,4
Schritt 6	0	53	7	88,3
	is learn 1	4	26	86,7
	Gesamtprozentsatz			87,8

a. Der Trennwert lautet ,500

Variablen in der Gleichung <sup>a</sup>						
		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.
Schritt 1 <sup>a</sup>	LEAA	,007	,002	20,489	1	,000
	Konstante	-2,264	,458	24,484	1	,000
	LEAA	,010	,002	21,831	1	,000
Schritt 2 <sup>b</sup>	QLAA	,016	,006	8,480	1	,004
	Konstante	-4,006	,875	20,961	1	,000
	LEAA	,010	,002	17,650	1	,000
Schritt 3 <sup>c</sup>	QLAA	,027	,009	9,533	1	,002

Schritt 4 <sup>d</sup>	QLIB	-,316	,155	4,171	1	,041	,729
	Konstante	-4,276	1,106	14,954	1	,000	,014
	LEAA	,012	,003	16,274	1	,000	1,012
	BEIB	-,657	,288	5,194	1	,023	,519
	QLAA	,039	,012	11,728	1	,001	1,040
Schritt 5 <sup>e</sup>	QLIB	-,428	,201	4,531	1	,033	,652
	Konstante	-5,269	1,396	14,253	1	,000	,005
	LEAA	,013	,003	13,668	1	,000	1,013
	BEIB	-,757	,336	5,087	1	,024	,469
	QLAA	,040	,013	9,447	1	,002	1,040
Schritt 6 <sup>f</sup>	QLIB	-,482	,205	5,551	1	,018	,617
	QLUES	,020	,008	5,586	1	,018	1,020
	Konstante	-6,531	1,799	13,184	1	,000	,001
	LEAA	,014	,004	12,321	1	,000	1,014
	BEIB	-,750	,354	4,488	1	,034	,473
Schritt 6 <sup>f</sup>	LEUES	,073	,055	1,781	1	,182	1,076
	QLAA	,041	,014	8,066	1	,005	1,042
	QLIB	-,518	,223	5,369	1	,020	,596
	QLUES	,024	,009	6,479	1	,011	1,024
	Konstante	-7,618	2,189	12,107	1	,001	,000

Tabelle 11-21 binär logistische Regression look up

Klassifizierungstabelle<sup>a</sup>

		Beobachtet	Vorhergesagt		
			is look up		Prozentsatz der Richtigen
			0	1	
Schritt 1	is look up	0	55	5	91,7
		1	2	28	93,3
	Gesamtprozentsatz				92,2
Schritt 2	is look up	0	57	3	95,0
		1	1	29	96,7
	Gesamtprozentsatz				95,6
Schritt 3	is look up	0	58	2	96,7
		1	0	30	100,0
	Gesamtprozentsatz				97,8
Schritt 4	is look up	0	58	2	96,7
		1	0	30	100,0
	Gesamtprozentsatz				97,8
Schritt 5	is look up	0	59	1	98,3
		1	1	29	96,7
	Gesamtprozentsatz				97,8

a. Der Trennwert lautet ,500

Variablen in der Gleichung

		RegressionskoeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1 <sup>a</sup>	LEAA	-,041	,011	14,241	1	,000	,960
	Konstante	2,231	,586	14,509	1	,000	9,308
Schritt 2 <sup>b</sup>	LEAA	-,052	,017	9,568	1	,002	,949
	LEEL	-,055	,026	4,599	1	,032	,947
	Konstante	3,672	1,011	13,184	1	,000	39,319
Schritt 3 <sup>c</sup>	BEBI	-,092	,072	1,646	1	,200	,912
	LEAA	-,062	,028	4,806	1	,028	,940
	LEEL	-,057	,054	1,118	1	,290	,944
	Konstante	4,913	1,619	9,205	1	,002	135,979
Schritt 4 <sup>c</sup>	BEBI	-,194	,082	5,562	1	,018	,824
	LEAA	-,066	,029	5,276	1	,022	,936
	Konstante	4,867	1,630	8,912	1	,003	129,916
Schritt 5 <sup>d</sup>	BEBI	-,224	,111	4,084	1	,043	,799
	LEAA	-,084	,040	4,399	1	,036	,919
	QLLK	-,047	,023	4,050	1	,044	,954
	Konstante	6,591	2,508	6,907	1	,009	728,420

Tabelle 11-22 binär logistische Regression casual - leisure

Klassifizierungstabelle<sup>a</sup>

		Beobachtet	Vorhergesagt		
			is browse		Prozentsatz der Richtigen
			0	1	
Schritt 1		0	57	3	95,0
	is browse	1	21	9	30,0
	Gesamtprozentsatz				73,3
Schritt 2		0	57	3	95,0
	is browse	1	18	12	40,0
	Gesamtprozentsatz				76,7
Schritt 3		0	57	3	95,0
	is browse	1	16	14	46,7
	Gesamtprozentsatz				78,9

a. Der Trennwert lautet ,500

Variablen in der Gleichung <sup>d</sup>							
		RegressionskoeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1 <sup>a</sup>	BEIG	,025	,010	6,407	1	,011	1,026
	Konstante	-1,124	,278	16,323	1	,000	,325
Schritt 2 <sup>b</sup>	BEIG	,025	,010	6,116	1	,013	1,025
	QLLK	,038	,015	6,600	1	,010	1,038
	Konstante	-1,521	,329	21,428	1	,000	,218
	BEIG	,026	,010	6,439	1	,011	1,026
Schritt 3 <sup>c</sup>	QLIG	1,192	1495,827	,000	1	,999	3,294
	QLLK	,041	,015	7,384	1	,007	1,042
	Konstante	-1,681	,348	23,278	1	,000	,186

Die Regressionsanalyse zeigt, dass die Nutzeraktionen als Prädiktoren für den Aufgabentyp unterschiedlich geeignet sind und vor allem für *casual-leisure* nicht den bisher entdeckten unterschiedlich verteilten Nutzeraktionen entsprechen (hier: BE-IG, QL-IG und QL-LK).

Bei den *look up* und *learn* Aufgaben dagegen zeigte sich wiederum die Lektüre von Text als geeigneter Prädiktor. Generell ist daraus zu schließen, dass die freie Suche erstens das unterschiedlichste Verhalten erzeugt und zweitens weiterhin die Zeit, die mit der Lektüre von Text verbracht wird, ein wichtiges Unterscheidungskriterium ist.

## 11.4 Wahl der Methoden

Die verwendeten Methoden werden im Folgenden kurz bewertet.

### 11.4.1 Eyetracker und Fragebögen

Die Verwendung des Eyetrackers und die Erstellung und Annotation der Videos mit den Labels war zwar zeitaufwändig, aber zielführend. Die Auswertung der 50 ms Frames bezüglich des jeweiligen Verhaltens des Nutzers war die Information, die zur Untersuchung der Frage nach der Informationsrezeption und der Präferenz für bestimmte Inhaltselemente in Abhängigkeit der Situation benötigt wurde.

Der automatisierte Ablauf und die Präsentation der Fragebögen auf dem Bildschirm funktionierten gut. Die Fragebögen konnten ohne Nachfragen zügig ausgefüllt werden (Eine Ausnahme bildete das teilweise fehlende Verständnis für die Kategorie *dominance* des SAM). Die Anzahl der Suchtasks war nicht zu hoch und konnte von den Probanden ohne erkennbar großen Aufwand ausgeführt werden.

In den Fragebögen wurden bei der Auswertung der Daten gezielte Fragen zu den Inhaltselementen vermisst (z.B. in welchen Inhaltselementen der Proband die gewünschte Information gefunden hat bzw. ob ggf. mehr Elemente eines bestimmten Typs gewünscht worden wären). Die Fragebögen und die Videolabels wurden deshalb für die zweite Studie überarbeitet.

#### 11.4.2 Elektromyograph

Die Verwendung des EMG war ausgesprochen aufwändig, die Anschaffung des Geräts und der Ersatz der Verschleißteile sind kostspielig. Auch für die Probanden war die Verwendung des EMG bisweilen unangenehm, da die Haut stark vorbehandelt werden musste.

Für den ursprünglich intendierten Zweck (Aufschluss über emotionale Reaktionen *während* der Suche, ohne den Probanden zu unterbrechen) lieferten die Daten des EMGs keine Hinweise. Ein Grund dafür war, dass die Messungen die Spannungszustände bestimmter Gesichtsmuskeln aufzeichnen, die über den langen Zeitraum der Suche stärker durch das Lesen und Betrachten der Seite als von affektiven Reaktionen beeinflusst wurden. Die durch möglicherweise vorhandene emotionale Reaktionen entstandene Bewegung der entsprechenden Muskelgruppen war nicht eindeutig von den weiteren Einflüssen abzugrenzen.

Michael Wittman [Wittman, 2012] zeigte allerdings einige *andere* interessante Aspekte, die aus den Daten des EMG erkennbar waren. So streuten die Aktivitäten der beobachteten Muskelgruppen bei den einzelnen Versuchspersonen generell sehr unterschiedlich.

Bei den mittels des EMGs erfassten Spannungswechseln der entsprechenden Muskelpartien, die auf emotionale Reaktionen hindeuten<sup>56</sup>, gab es allerdings keine statistisch signifikanten Unterschiede zu beobachten. Der Zusammenhang mit den Angaben zur Selbsteinschätzung der emotionalen Lage mittels des *Self Assessment Manikin* war relativ schwach, so [Wittman, 2012].

---

<sup>56</sup> Über alle Versuchspersonen und über alle Aufgaben.

Allerdings zeigten sich Unterschiede bei den Reaktionen der einzelnen Testpersonen *zwischen* den verschiedenen Aufgaben *einer* Person [Wittman, 2012]. Wittman zeigt weiter, dass viele lokale Maxima bei der Anspannung des *Corrugator* vorhanden waren, die jeweils weit höher lagen als die Verteilung der Anspannung außerhalb der Maxima. Bei über 50% dieser stark erhöhten Spannungszustände navigierte der Nutzer gerade auf eine neue Seite und musste sich mit neuen Inhalten befassen.

In der hier durchgeführten Folgestudie wurde darauf verzichtet, das EMG erneut zu verwenden, da der für die Erkenntnisse dieser Arbeit notwendige Nutzen ausblieb und somit der hohe Aufwand bei der Verwendung des Gerätes nicht gerechtfertigt war.

#### 11.4.3 Self Assessment Manikin

Die Verwendung des *Self Assessment Manikin* [Bradley und Lang, 1994] erwies sich bis auf die genannten Probleme als für die Probanden gut handhabbar.

Die Interpretation der Angaben zur Selbsteinschätzung mittels der einzelnen Darstellungen bei der Auswertung ist schwierig. Grundsätzlich ist zwar eine Reihenfolge bei der Abfolge der je fünf Darstellungen vorhanden im Sinne einer Ordinalskala, allerdings bedeutet das nicht zwingend, dass diese Skalen in Kombination *nur* von positiv zu negativ bzw. umgekehrt verlaufen. Es können bestimmte Kombinationen der Ausprägungen (leichter Erregung und zusätzlich stark ausgeprägte angenehme Emotionen) vorhanden sein, die möglicherweise subjektiv als „positiver“ empfunden werden als z.B. die selben Ausprägungen der angenehmen Emotionen in Kombination mit *keiner* Erregung. Das bedeutet, dass die Annahme, die hier bei der quantitativen Auswertung des SAM getroffen wurde, dass die Punkte der beide Dimensionen unabhängig voneinander vergeben wurden, fehlerbehaftet sein kann.

Zusätzlich zeigte sich bei der Verwendung des SAM im Experiment, dass die SD der Angaben relativ gering ist, genauso die Varianz. Der Mittelwert ist bei beiden hier ausgewerteten Dimensionen um die „zweitpositivste“ Ausprägung angesiedelt. Die Verteilung der Angaben ist relativ homogen, sie sind mit einer Irrtumswahrscheinlichkeit von 0,000 % nicht normalverteilt (siehe Tabelle 11-23).



Da hier aber eher ein Stichproben-Problem<sup>57</sup> denn ein methodisches vermutet wird, wird diese Methode zur Selbstauskunft der emotionalen Lage der Probanden in der zweiten empirischen Studie weiter verwendet werden. Auch wenn in der Laborsituation und/oder auch durch Zufall keine stark unterschiedlichen emotionalen Zustände auftraten, so ist es dennoch wichtig, die Stimmung der Probanden zu erfassen, da nicht *grundsätzlich* von einer homogenen emotionalen Lage ausgegangen werden kann.

Tabelle 11-23 Test auf Normalverteilung der Angaben in den Dimensionen arousal und valence

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
SAM_A	,219	89	,000	,874	89	,000
SAM-P	,298	89	,000	,833	89	,000

a. Signifikanzkorrektur nach Lilliefors

In den Experimenten entstanden unterschiedliche Typen empirischer Daten. Der Vorteil der maschinell entstanden Logs (Browser-Interaktion und Blickbewegungen) ist, dass die Daten objektiv sind, sie haben allerdings den Nachteil, dass die Menge der entstandenen Daten relativ groß ist und ohne Vorverarbeitung keine intuitiven Aussagen über diese Daten möglich sind.

Die Fragebögen wurden von den Nutzern selbst ausgefüllt, dadurch enthalten sie subjektive Angaben der Testpersonen. Ein Vorteil dabei ist, dass eine Beeinflussung des Versuchsleiters (durch Fragen oder Nachfragen) ausgeschlossen werden konnte, einen gewissen Nachteil allerdings stellt die Tatsache dar, dass unklar bleibt, *was genau* die jeweilige Testperson tatsächlich unter der Frage verstanden hat.

In der Summe sind die Daten eine Mischung zwischen objektiven und subjektiv erfassten Daten im Sinne der Kombination von explizitem und implizitem Feedback.

#### 11.4.4 Browserlogs

Die Browserlogs wurden nur teilweise ausgewertet [siehe dazu Knäusl et al., 2012]. Eine Vorhersage der Videolabels anhand der Browser-Interaktion war nicht möglich. Dazu müssen genauere Untersuchungen der Log-Dateien vorgenommen werden wie

<sup>57</sup> Im Sinne einer zufällig sehr homogenen emotionalen Ausgangsstimmung.

---

eine zusätzlich Cluster-Analyse der unterschiedlichen Interaktionstypen (siehe dazu auch: [Guo und Agichtein2010-2]).

## 12 Zweite Studie zum Kontextabhängigen Nutzerverhalten

Die erste Studie zeigte, dass Zusammenhänge zwischen den Nutzerszenarien und den Präferenzen für bestimmte Inhaltselemente nachweisbar sind. Um diese Ergebnisse zu überprüfen und darüber hinaus Erkenntnisse und Verbesserungen in Designfragen einbringen zu können, wurde erneut eine empirische Erhebung durchgeführt.

Vor allem der Wunsch der Nutzer nach *zusätzlichen* Inhaltselementen wurde in der ersten Studie nicht berücksichtigt. Da aber die Möglichkeit besteht, dass die Nutzer zwar alle von einem Typ verfügbaren Elemente in ihre Suche miteinbeziehen, diese aber dennoch nicht als ausreichend bewerten, muss zusätzlich zur Interaktion auch nach darüber hinaus gewünschten Elementen gefragt werden. Das war ein wichtiger Aspekt für die Durchführung einer zweiten Studie.

Eine zweite Schwäche der ersten Studie stellte teilweise die Benennung der Videolabels dar. Die Auswertungen zeigten, dass das Scannen und Lesen einer Liste ein Merkmal zur Unterscheidung der Nutzungsszenarien ist. Es wurde aber sowohl für das Inhaltsverzeichnis als auch für alle anderen Listen in den Artikeln nur die Bezeichnung LI verwendet, so dass eine Unterscheidung nur anhand der Videos selbst, nicht aber durch die Labels möglich war. Dieses Problem wurde in der anschließenden Studie behoben.

### 12.1 Aufbau und Ablauf

In dieser Studie wird wiederum die situative und subjektive Bewertung der Inhalte, die die Informationsrezeption der Nutzer erfasst, untersucht:

- 1.) Durch die Eyetracking Software werden wieder Videos erstellt, entsprechend Studie eins.
- 2.) Mit Hilfe von Fragebögen wird nach jedem Task die subjektive Zufriedenheit mit der Erfüllung der Aufgabe abgefragt. Dabei liegt der Fokus vor allem auf der Erfüllung des Nutzerbedürfnisses durch die einzelnen vorhan-

denen Inhaltselemente. Deshalb wird gezielter abgefragt, welche Inhaltselemente relevant waren und für welchen Typ mehr Bedarf vorhanden gewesen ist.

Für die vollständigen Fragebögen und die Tasks siehe Anhang 6 und 7.1 – 7.3. Laboraufbau und weitere Methoden entsprechen der ersten Studie.

Wie auch die erste Studie fand die Erhebung der Daten an der Universität Regensburg im Eyetracking-Labor statt, das nach einem Umzug nun aus nur noch einem Raum besteht. Der Stimulus-PC und der Eyetracker befanden sich deshalb hinter einer Stellwand, so dass die Versuchsperson keinen Sichtkontakt zum Versuchsleiter und zum Test-PC hatte. Der PC mit der iView-Software und ein zweiter Monitor, der den Bildschirm des Stimulus-PC übertrug, standen auf der Seite des Versuchsleiters.

Fragebögen, Testabläuferläuterung und Suchaufgaben wurden wieder automatisch auf dem Stimulus-PC präsentiert, so dass der Versuchsleiter nur im Falle eines Problems eingreifen musste und ansonsten die Testperson das Experiment selbst steuerte.

Die Bearbeitungszeit für die Aufgaben wurde wieder begrenzt (siehe dazu auch 11.2.3), der Nutzer konnte die Aufgaben allerdings auch selbst beenden, wenn er den Task als erledigt erachtete (maximal verfügbare Zeit *look-up*: zwei Minuten, *learn*: vier Minuten und *casual-leisure*: fünf Minuten). Die Tasks wurden wieder randomisiert präsentiert, um einen Lerneffekt zu vermeiden (vgl. dazu [Bortz,2001])

Jede VP wurde gebeten, sechs Tasks abzuarbeiten. Das Vorgehen bei der Auswahl der Themen, die Präsentation der Aufgaben und die verwendete Soft- und Hardware entsprechen jeweils der ersten Erhebung.

## 12.2 Verarbeitung der Daten

### **Videolabels**

Die Videolabels wurden nach den Erkenntnissen aus der ersten Studie überarbeitet und es wurden englische Bezeichnungen eingeführt.

Tabelle 12-1 Videolabels "Inhaltselement"

Label-Name	Inhaltselement
<b>HL</b>	Headline
<b>TX</b>	Text [Artikel-Absatz]
<b>LI</b>	List
<b>IX</b>	Index
<b>IN</b>	Introduction
<b>PI</b>	Picture (Einordnung erfolgt nach Klassifikation Schubart/Knäusl)
<b>IG</b>	Infografik (Einordnung erfolgt nach Klassifikation Schubart/Knäusl)
<b>IB</b>	Infobox (proprietäres Wikipedia-Element)
<b>LINK</b>	Link (incl. der Verweise am Ende der Wikipedia-Seite)

Tabelle 12-2 Mögliche Kombinationen der Videolabels

Aktions-Typen	Inhalts-Element-Typen
<b>Read [RE]</b>	Text [TX] Infografik [IG] List [LI] Index [IX] Infobox [IB] Introduction [IN] Links [LINK] Headline [HL]
<b>Scan [SC]</b>	Text [TX] Headline [HL] Infografik [IG] Liste [LI] Picture [PI] Infobox [IB] Introduction [IN] Index [IX] Links [LI]
<b>Examine [EX]</b>	Infografik [IG] Picture [PI]
<b>Navigieren [NA]</b>	Navigieren auch: Scrollen, Suche mit Strg. F

Die Videos wurden wieder manuell mit den beschriebenen Labels annotiert. Beim vorliegenden Versuchsaufbau gab es keine Alternative zu diesem Vorgehen. Es gibt, wie bereits in der Auswertung der Daten der ersten Studie, zu jedem Zeitpunkt zwei Labels, davon eines die Aktion des Nutzers betreffend und eines mit dem entsprechenden Inhaltselement. Um das Vorgehen beim Annotieren der Videos transparenter zu machen seien noch folgende Hinweise gegeben:

Die Vergabe der Aktionslabels unterliegt der Interpretation des Blickverhaltens der Testperson auf Basis der Literatur [Duchowski, 2007] und den Erfahrungswerten aus der ersten Studie.

Betrachtet der Nutzer ein Bild oder eine Infografik, so ruht das Auge entweder verlängert auf einem Punkt oder bewegt sich mit kurzen Sakkaden und kurzen Fixationen innerhalb einer bestimmten Fläche (dem betrachteten Bild oder der Infografik, Aktion: *examine*). Das Aktionslabel NA für navigieren wurde immer dann vergeben, wenn der Nutzer entweder eine Query eingab, einem Link folgte oder über den Bildschirm scrollte, *ohne* dabei Inhalte zu fixieren.

Lange Navigationszeiten vor allem zu Beginn des Tasks entstanden hauptsächlich durch Unsicherheiten bei Formulierung der Query – entsprechend [Belkin, 1980]<sup>58</sup>. In der Regel wurde jedem Aktionslabel immer ein Inhaltselement zugeordnet, Ausnahmen sind dabei NA, das *immer* ohne Inhaltselement verwendet wird und SC für Scan, das *unter Umständen* ohne Inhaltselement verwendet wird, da in manchen Fällen der Blick zeitweise ziellos die Seite scannte und kein konkretes Inhaltselement Gegenstand von Fixationen war.

Die erste Studie zeigte, dass bei den drei Aufgabentypen vor allem ein Unterschied bei der Informationsrezeption besteht zwischen den Anteilen der Nutzeraktionen, die der Orientierung bzw. Navigation dienen im Vergleich zu denen zum Zwecke der Informationsaufnahme bei den drei Aufgabentypen (siehe dazu ab Kapitel 11.3.4.). Um deshalb die Verteilung der navigierenden und informierenden Aktionen bei den verschiedenen Aufgabentypen zu untersuchen, werden hier weiter die Nutzeraktionen (Video-

---

<sup>58</sup> In den Videos erkennbar.

label-Kombinationen) in *navigierende* und *informierende* Klassen unterteilt [vgl. dazu auch Broder, 2002].

*Tabelle 12-3 Unterscheidung in navigierende und informierende Labelkombinationen*

<b>Navigierend [NA]</b>	<b>Informierend [IN]</b>
Scan – Headlines	Read – Text
Read – Headlines	Examine - Picture
Scan - Index	Read – Infobox
Read – Index	Read – List
Navigate	Examine - Infografik
Scan	
Scan - Text	
Read – Link	
Scan – Link	
Scan - List	
Scan – Picture/Infografik/Infobox	

Die Verteilung der navigierenden und informierenden Aktionen wurde auf Gleichverteilung über die drei Tasktypen (= V1, siehe Abbildung 12-1) untersucht. Überprüft wurden hier die Auftretenswahrscheinlichkeiten (Mittelwerte der relativen Häufigkeiten, siehe dazu auch Kapitel 0) der Videolabel-Kombinationen, die entsprechend den Klassen NA oder IN zugeordnet wurden. Der Kruskal-Wallis-Test zeigt, dass die Verteilung über die drei Aufgabentypen nicht gleich ist (101 Tasks,  $n_{\text{learn}} = 37$ ,  $n_{\text{look up}} = 32$ ,  $n_{\text{browse}} = 32$ ).

**Übersicht über Hypothesentest**

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von IN ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
2	Die Verteilung von NA ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

Abbildung 12-1 Kruskal-Wallis-Test über die Verteilung von NA und IN über alle drei Tasktypen

In den verschiedenen Szenarien (=V1) zeigen die Nutzer unterschiedliches Verhalten in den Bereichen der Informationsaufnahme (informieren) und der Orientierung (navigieren). Die beiden Aktions-Klassen bzw. die dadurch entstehende Informationsrezeption sind in den Szenarien *nicht* gleichverteilt.

Noch deutlicher werden die Ergebnisse, setzt man die jeweiligen relativen Anteile von NA-Aktionen und IN-Aktionen ins Verhältnis. Dann erhält man folgende Aufstellung *casual-leisure*: IN:NA  $\approx$  62:38, *look-up*: IN:NA  $\approx$  37:62 und *learn* IN:NA  $\approx$  72:28

Tabelle 12-4 Verhältnis von IN zu NA bei allen drei Tasktypen

	CL-IN	CL-NA	LU-IN	CL-NA	LE-IN	LE-NA
Mittelwert Wahrscheinlichkeit alle Tasks eines Typs	p(IN) = 0,55684	p(NA) = 0,34316	p(IN)= 0,33472	p(NA)= 0,56528	p(IN)= 0,63363	p(NA) = 0,36637
Relativ (%) Summe aller Tasks eines Typs	0,6187	0,3813	0,3719	0,6281	0,7182	0,2818



Bei den *look up* Aufgaben wurde am meisten navigiert, bei *learn* am wenigsten. Da die *casual-leisure* Szenarien der zweiten Studie nicht optimal gestellt waren<sup>59</sup> wurde auch hier eher mehr navigiert, wenngleich auch deutlich weniger als bei *look up*. Hier steht das Auffinden einer endlichen, klar definierten Information im Vordergrund. Der Nutzer geht also dabei stärker selektiv vor als bei den *learn* Szenarios. Der deutlich kleinere Anteil an navigierenden Aktionen bei *learn* zeigt, dass der Nutzer durch das weniger stark definierte und endliche Ziel auch weniger selektiv vorgeht und im vorliegenden Fall die meiste Zeit des Tasks tatsächlich mit der Informationsaufnahme beschäftigt war. Vor dem Hintergrund des Ziels einer adaptiven Informationspräsentation ist das insofern relevant, als dass – wenn die navigierenden von den informierenden Tätigkeiten unterschieden werden – durch diese Interaktion des Nutzers mit den Inhalten Rückschlüsse auf sein inhaltliches Suchziel (im Sinne von  $S_1$  und  $S_2$ , siehe dazu Kapitel 6.2.1) gezogen werden können.

## 12.3 Auswertung der Daten

Die gewonnenen Daten werden hinsichtlich der Fragestellung, ob die Nutzer abhängig vom Nutzungsszenario bei der Suche in der Wikipedia unterschiedliche Inhalte bevorzugen, ausgewertet.

Die Datensätze der Videos, die durch den Eyetracker erstanden, wurden entsprechend der ersten Studie verarbeitet (siehe dazu auch 11.3.2)

### 12.3.1 Demografische Angaben

Die Probanden sind zwischen 1983 und 1994 geboren. Neun der Versuchspersonen sind weiblich, 19 männlich. 20 der 28 Probanden beantworteten die Frage, ob sie die Wikipedia auch als Gegenstand ihrer Freizeitbeschäftigung verwenden, mit drei oder mehr Punkte (Durchschnitt gesamt: 3,46 Punkte, verwendet wurde auch hier, so weit nicht anders angegeben, in den Fragebögen eine Likertskala von 0-5 in Einer-Schritten).

---

<sup>59</sup> Die VPs sollten die Funktion „Zufälliger Artikel“ verwenden, um von dort aus zu persönlich interessanten Themen weiter zu navigieren. Allerdings kehrten die meisten VPs immer wieder zur Funktion „zufälliger Artikel“ zurück.

Die Einschätzung der Seriosität der Wikipedia wurde im Durchschnitt mit 2,96 Punkten angegeben. Für die Auswertungen konnten in der zweiten Studie 101 Tasks ( $n_{\text{learn}} = 37$ ,  $n_{\text{look up}} = 32$ ,  $n_{\text{browse}} = 32$ ) verwendet werden.

### 12.3.2 Auswertung des emotionalen Zustandes der Probanden nach jedem Task

Die Nutzer wurden im *Post-Task*-Fragebogen (Fragebogen nach jeder Aufgabe, siehe Anhang 7.3) gebeten, mit Hilfe des Self-Assessment-Maniking [Bradley und Lang, 1994] Auskunft über ihre momentane emotionale Stimmung zu geben. Da in der ersten Studie verwendet die Testpersonen teilweise die dritte Dimension des SAM, *Dominance*, nicht gut einordnen konnten wurde diese in der zweiten Studie weggelassen, so dass sich die Angaben auf den grad der Anspannung und der Erfreutheit der Probanden beschränken.

Die Probanden kreuzten jeweils das momentan für sie am besten zutreffende „Männchen“ (eine von fünf möglichen Ausprägungen des SAM) an ( $n = 168$ ). Die Auswertung der Bilder erfolgte entsprechend der ersten Studie (für Bild 1. in der Dimension *valence* fünf Punkte absteigend bis Bild 5., das mit einem Punkt bewertet wurde. Bei der zweiten Dimension *arousal* erfolgte die Punktevergabe umgekehrt: Bild 1 wurde mit einem Punkt bewertet, aufsteigend bis Bild 5, für das fünf Punkte vergeben wurden, siehe dazu auch 11.3.8 und Tabelle 11-17).

Personen, die in der Dimension *valence* nur zwei oder einen Punkt angegeben haben werden in die Kategorie „negative Stimmung“ eingeteilt. Gleiches gilt für die Dimension *arousal*, Personen mit zwei oder einem Punkt bilden die Gruppe mit eher negativer Stimmung (hohe Anspannung, siehe dazu auch Tabelle 11-17).

Davon ausgehend, dass eine Person eine eher negative emotionale Lage hat, wenn sie in mindestens einer der beiden Bewertungen unter dem Grenzwert für die „positive“ Stimmung liegt, war bei 18 von 28 Probanden bzw. bei insgesamt 36 Aufgaben (21 mal negativ in der Dimension *valence*, 15 mal negativ in der Dimension *arousal* bzw. drei mal in beiden Dimensionen) mindestens einmal die emotionale Lage eher negativ. Bei 111 Aufgaben und insgesamt 25 Probanden wurde die Dimension *arousal* im positiven Bereich angegeben, also Abwesenheit von Anspannung. Numerisch wurden hier alle Werte größer als drei berücksichtigt. In 90 Fällen und bei insgesamt 26 Probanden ga-

ben die Testpersonen bei der Dimension *valence* einen Status an, der mit höher als drei bewertet wurde. 59 der beschriebenen Fälle liegen sowohl in der positiven Klasse der ersten also auch der zweiten Dimension und somit ergibt das in Summe 142 Fälle, in denen bei den Probanden eine tendenziell positive Stimmung herrschte.

Die Histogramme in Abbildung 12-2 zeigen jeweils die Häufigkeiten der Werte in beiden Dimensionen. Auch hier kommen die „negativen“ Ausprägungen (also Werte kleiner oder gleich zwei) wie bereits in der ersten Studie weniger oft vor. Die Mittelwerte liegen bei der Dimension *valence* bei 3,51 und Dimension *arousal* bei 3,63. Das 95%-Konfidenzintervall des Mittelwerts liegt bei der Dimension *arousal* zwischen 3,37 und 3,66 und bei *valence* zwischen 3,65 und 3,94. Die Streuung ist also wiederum relativ gering.

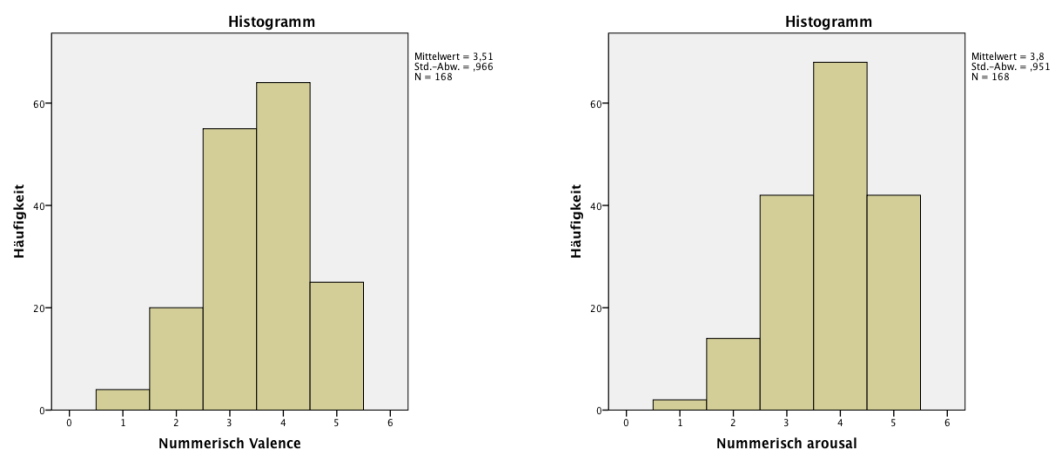


Abbildung 12-2 Histogramm der Angaben zur Selbsteinschätzung der emotionalen Lage mittels SAM über alle tasks (valence 1=angenehm, 5= unangenehm, arousal 1= erregt, 5 = ruhig)

Tabelle 12-5 Test auf Normalverteilung der Angaben mittels SAM

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Valence	,201	101	,000	,900	101	,000
Arousal	,254	101	,000	,823	101	,000

a. Signifikanzkorrektur nach Lilliefors

Die Angaben zur emotionalen Selbstauskunft sind nicht normalverteilt (siehe Tabelle 12-5). Die Auswertungen bestätigen den Trend der ersten Studie zu einem relativ ho-

mogenen Feld der emotionalen Lage bei den Probanden. Im Umkehrschluss zu den Überlegungen aus Kapitel 5.6.7, in dem der Einfluss von Emotionen auf das Suchverhalten und das allgemeine Verhalten der Menschen dargelegt wird, ist hier davon auszugehen, dass die Probanden kaum oder nicht *unterschiedlich* durch ihre emotionale Lage beeinflusst wurden (da diese relativ homogen war). Um dies weiter zu überprüfen wurden auch die Angaben in den Fragebögen auf einen Zusammenhang mit der emotionalen Selbstauskunft untersucht.

Im *Post-Task* Fragebogen wurden die Probanden gefragt, ob sie sich für die Erledigung der jeweiligen Aufgabe mehr Inhalte eines bestimmten Typs gewünscht hätten ( $n = 168$ ). Es wurde nun der Einfluss der dort angegebenen emotionalen Lage auf den Wunsch nach bestimmten Inhaltstypen überprüft (Emotion als Einflussgröße). Die Probanden wurden dazu in zwei Gruppen eingeteilt (negative und positive Stimmung) da wie beschrieben hier die Streuung relativ gering war und somit die Gruppenunterschiede hinsichtlich dieses Merkmals bei zwei Gruppen am stärksten ausfallen (siehe Tabelle 12-6). Es wurde der t-Test für den Vergleich der Mittelwerte zweier Stichproben verwendet, obwohl dieser eigentlich eine Normalverteilung der Variablen voraussetzt, da die Gruppengrößen der verglichenen Stichproben hier hinreichend groß sind ( $n_{\text{positiv}} = 110$  und  $n_{\text{negativ}} = 58$ ), siehe dazu [Leonhart, 2013].

Tabelle 12-6 Lageparameter und Häufigkeiten valence und arousal

Statistiken		Nummerisch Valence	Nummerisch arousal
N	Gültig	168	168
	Fehlend	10	10
Mittelwert		3,512	3,797
Standardabweichung		,9636	,9485
Varianz		,928	,900

**Häufigkeitstabelle**

Numerisch Valence				
	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
	1,0	4	2,2	2,4
	2,0	20	11,2	14,2
	3,0	55	30,7	46,7
Gültig	3,5	1	,6	47,3
	4,0	64	35,8	85,2
	5,0	25	14,0	100,0
	Gesamt	169	94,4	100,0
Fehlend	System	10	5,6	
Gesamt		179	100,0	

Numerisch arousal				
	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
	1,0	2	1,1	1,2
	2,0	14	7,8	9,5
	3,0	42	23,5	34,3
Gültig	3,6	1	,6	34,9
	4,0	68	38,0	75,1
	5,0	42	23,5	100,0
	Gesamt	169	94,4	100,0
Fehlend	System	10	5,6	
Gesamt		179	100,0	

Hinsichtlich der Angaben nach dem Wunsch nach bestimmten Inhaltstypen unterscheiden sich die beiden Gruppen (in SPSS Gruppe 1 Dimension *arousal*  $\geq 4$  ( $n = 110$ ), Gruppe 2 Dimension *arousal*  $< 4$  ( $n=58$ )) bei dem Wunsch nach mehr Fotos.

Dieser Wert besagt, dass die Personen, der Gruppen positiv oder negativ in der Selbsteinschätzung der Dimension *arousal*, einen statistisch signifikant unterschiedlichen Wunsch nach mehr Fotos geäußert haben ( $p = 0,003$ , max. Irrtumswahrscheinlichkeit für Test mit drei Variablen  $10/3$ , also  $3,33\%$  oder  $0,033$ ). Das ist ein Hinweis auf die Emotionen als Einflussgröße. Weiterhin zeigt dieses Ergebnis auch, dass, wie in Kapitel 1 besprochen wurde, Bilder aus unterschiedlichen persönlichen Gründen verwendet bzw. berücksichtigt werden.

Tabelle 12-7 t-Test "Mehr Foto/Text/Grafiken gewünscht" der Gruppen positiv "arousal" und negativ „arousal“ (Trennwert 4)

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. [2-seitig]	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
mehr Fotos gewünscht	Varianzen sind gleich	,002	,964	-	166	,003	-,5787	,1915	-,9568	-,2005
	Varianzen sind nicht gleich			-	117,875	,003	-,5787	,1905	-,9559	-,2015
mehr Graphiken/Tabellen gewünscht	Varianzen sind gleich	1,687	,196	-,812	166	,418	-,1652	,2034	-,5668	,2364
	Varianzen sind nicht gleich			-,826	121,603	,411	-,1652	,2001	-,5613	,2308
mehr Text gewünscht	Varianzen sind gleich	,561	,455	-,977	166	,330	-,1853	,1896	-,5596	,1891
	Varianzen sind nicht gleich			-,995	122,239	,322	-,1853	,1861	-,5538	,1832

Tabelle 12-8 t-Test "Mehr Foto/Text/Grafiken gewünscht" der Gruppen positiv "valence" und negativ "valence" (Trennwert 4)

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. [2-seitig]	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
mehr Fotos gewünscht	Varianzen sind gleich	2,565	,111	2,030	166	,044	-,3758	,1851	-,7412	-,0103
	Varianzen sind nicht gleich			2,020	159,548	,045	-,3758	,1860	-,7431	-,0084
mehr Graphiken/Tabellen gewünscht	Varianzen sind gleich	9,333	,003	3,556	166	,000	,6655	,1871	,2960	1,0350
	Varianzen sind nicht gleich			3,516	151,729	,001	,6655	,1893	,2915	1,0394
mehr Text gewünscht	Varianzen sind gleich	,020	,888	-,252	166	,801	-,0457	,1811	-,4032	,3119
	Varianzen sind nicht gleich			-,253	164,646	,801	-,0457	,1808	-,4026	,3113

Anders verhält es sich bei der Bevorzugung von Text und/oder Bild laut Fragebogen für die Gruppen positiv und negativ nach Tabelle 12-8, hinsichtlich der Dimension „valence“ (auch hier Trennwert 4,  $n_{\text{positiv}} = 89$ ,  $n_{\text{negativ}} = 79$ ). Hier zeigt die Untersuchung der Mittelwerte, dass sich zusätzlich die Angaben zum Wunsch nach mehr Grafiken/Tabellen und wiederum die zum Wunsch nach Bildern zwischen den Gruppen negativ und positiv im Bereich *valence* statistisch signifikant unterscheiden.

Die Auswertung der Videolabels hinsichtlich eines Zusammenhangs mit der emotionalen Selbsteinschätzung der Probanden und deren Interaktion zeigt, dass die Gruppen *arousal\_positiv* ( $\geq 4$  Punkte in der SAM-Kategorie *arousal*,  $n_{\text{positiv}} = 77$ , und *arousal\_negativ* ( $< 4$  Punkte in der SAM-Kategorie *arousal*,  $n_{\text{negativ}} = 24$ )<sup>60</sup> sich ebenfalls bei der Berücksichtigung von Bildern und zusätzlich bei der Lektüre von Text schwach signifikant unterscheiden (siehe t-Test im Anhang 8.7,  $p = 0,012$  bei RE-TX und  $p = 0,018$  bei EX-PI).

Bei einer Untersuchung der zweiten Kategorie des SAM, die den Grad der Anspannung des Probanden widerspiegelt, wurden ebenfalls Unterschiede in der Suchstrategie und der Verwendung bestimmter Inhaltselemente deutlich. Hier zeigt der t-Test (Trennwert der beiden Gruppen war wiederum 4,  $n_{\text{positiv}} = 48$ ,  $n_{\text{negativ}} = 53$ ), dass Unterschiede bei der Lektüre des Inhaltsverzeichnis, dem Lesen der Überschriften und allgemein der Zeit, die mit navigierenden Tätigkeiten verbracht wurde, bestehen (siehe Anhang 8.7).

Diese Erkenntnisse bieten auch eine Grundlage für Rückschlüsse auf die emotionale Lage des Nutzers aus der Interaktion (siehe dazu auch Kapitel 7.2.3). Im Umfang der navigierenden und überblicksartigen Aktionen finden sich nach den vorliegenden Daten Hinweise auf seinen Grad der Anspannung.

### 12.3.3 Präferenz für Inhaltselemente im Zusammenhang mit dem Tasktyp

In den *Post-Task*-Fragebögen wurden die Probanden nach jeder Aufgabe dazu befragt, ob sie sich für die Lösung der Aufgabe mehr Text, und/oder mehr Bilder und/oder mehr Infografiken gewünscht haben. Dies diene als Ergänzung zur Beobachtung der Interaktion, denn selbst wenn ein Proband z.B. alle vorhandenen Bilder betrachtet hat, so kann dennoch darüber hinaus der Wunsch nach mehr Bildern bestehen. Die Mittelwertgleichheit der Angaben zu den Wünschen nach bestimmten Inhaltselementen wurden mittels des Tamhane-T2<sup>61</sup> (Gruppenstärke je  $n = 56$ , drei Gruppen = Aufgabentypen) untersucht, um zu überprüfen, ob beim Wunsch nach bestimmten Inhaltselementen statistisch signifikante Unterschiede zwischen den verschiedenen Nutzungsszenarien vorliegen. Das alpha-Fehler Niveau wird wieder entsprechend angepasst: 10/3,

---

<sup>60</sup> Die Gruppengrößen variieren je nachdem ob „nur“ die Fragebögen ausgewertet wurden ( $n = 168$ ) oder zusätzlich die Nutzeraktionen aus den Videos (hier ist  $n = 101$ ).

<sup>61</sup> Da bei den Stichproben keine Varianzhomogenität zwischen den Gruppen und keine Normalverteilung vorliegen, siehe [Leonhart, 2013].

also eine max. 3,33% Irrtumswahrscheinlichkeit ( $p = 0,033$ ) wird akzeptiert, um  $H_0$  zu verwerfen.

Tabelle 12-9 Mehrfachvergleich der Aufgabentypen auf unterschiedliche Wünsche

Mehrfachvergleiche							
Tamhane							
Abhängige Variable	(I) V3	(J) V3	Mittlere Diffe- renz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
mehr Text gewünscht	1	2	-,4036	,2358	,247	-,976	,169
		3	-,4717	,2191	,099	-1,005	,061
	2	1	,4036	,2358	,247	-,169	,976
		3	-,0681	,1905	,978	-,531	,394
	3	1	,4717	,2191	,099	-,061	1,005
		2	,0681	,1905	,978	-,394	,531
mehr Fotos gewünscht	1	2	,6066	,2266	<b>,026</b>	,057	1,157
		3	,6415	,2131	<b>,010</b>	,124	1,159
	2	1	-,6066	,2266	<b>,026</b>	-1,157	-,057
		3	,0349	,2309	,998	-,526	,595
	3	1	-,6415	,2131	<b>,010</b>	-1,159	-,124
		2	-,0349	,2309	,998	-,595	,526
mehr Graphi- ken/Tabellen gewünscht	1	2	-1,0367	,2362	<b>,000</b>	-1,611	-,463
		3	-,4717	,2381	,144	-1,050	,107
	2	1	1,0367	,2362	<b>,000</b>	,463	1,611
		3	,5650	,2016	<b>,018</b>	,076	1,054
	3	1	,4717	,2381	,144	-,107	1,050
		2	-,5650	,2016	<b>,018</b>	-1,054	-,076
*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.							

Der Test für die Gruppen *learn* (in der Tabelle als 2 codiert) und *casual-leisure* (in der Tabelle als 3 codiert) ergab, dass lediglich der Wunsch nach mehr Graphiken und Tabellen unterschiedlich ausfiel (siehe Tabelle 12-9).

Anders verhält es sich bei dem Vergleich der Mittelwerte zwischen den Gruppen *casual-leisure* und *look up* (in der Tabelle codiert als 3 und 1). Bei diesen beiden Tasktypen wurde der Wunsch nach mehr Elementen eines Typs bei Grafiken und Fotos abhängig vom Tasktyp unterschiedlich stark angegeben. Dieser Unterschied besteht auch bei Verwendung der Inhalte (Videolabels) bei der Problemlösung.

Beim Vergleich der Daten zwischen den Aufgabentypen *learn* und *look up* zeigte der Test, dass der Wunsch nach Graphiken/Tabellen und wiederum bei Bildern unterschiedlich stark ausgeprägt war.



Es wurde bei dieser Auswertung allerdings nicht berücksichtigt, dass es weitere Faktoren gibt, die vor allem den Wunsch nach Bildern beeinflussen (siehe Kapitel 10.5). Bei der Erstellung der Aufgaben für die Datenerhebung wurde zwar durch einen Vergleich sichergestellt, dass für alle bearbeiteten Themen jeweils etwa gleichviele Bilder zu Verfügung standen, letztendlich aber determinierte der konkrete Navigationspfad des Nutzers die verfügbare Anzahl der Bilder. Darüber hinaus kann auch das Thema selbst möglicherweise [siehe dazu auch Vogel, 2013] die Präferenzen für Bilder beeinflussen. Ebenfalls nicht von der Hand zu weisen ist die Möglichkeit, dass z.B. bestimmte Arten von Bildern [Knäusl et al., 2012-1; Schubart, 2013; Knäusl und Schubart, 2013 und Kapitel 10.7] vermisst wurden. Es ist möglich, dass objektiv quantitativ die Bilder zwar gleichmäßig verteilt waren, nicht aber hinsichtlich z.B. ihrer subjektiv empfundenen emotionalen, inhaltlichen und/oder technischen Qualität.

Die Fragen nach den Wünschen der Nutzer geben neben der Beobachtung des Interaktionsverhaltens vor allem Aufschluss darüber, was der Nutzer in Abhängigkeit von seiner Aufgabe subjektiv präferiert. Bei der Erhebung wurde gezielt im Zusammenhang mit der konkreten Durchführung einer Suchaufgabe nach den Wünschen gefragt, um den Nutzer in einer authentischen Situation Selbstauskunft geben zu lassen (im Gegensatz zu einer rein theoretischen Befragung, welche Wünsche er bei den verschiedenen Aufgaben äußern *würde*).

#### 12.3.4 Auffinden der gesuchten Information

Es wurde weiter der Frage nachgegangen, in welchen Inhaltselementen die Probanden jeweils die gewünschte Information gefunden haben (Gruppenstärke je  $n = 56$ , drei Gruppen = Aufgabentypen). Vor allem bei der Aufgabe L2 war die gesuchte Information sowohl dem Text als auch einer Infografik zu entnehmen, bei den *learn* und *casual leisure* Aufgaben theoretisch in jedem Info-Element.

Die statistische Auswertung zeigt Unterschiede (siehe Tabelle 12-10). Bei den Typen *casual-leisure* und *look up* unterscheiden sich die Mittelwerte statistisch signifikant (max. erlaubte Irrtumswahrscheinlichkeit 3,3%) für die Bereiche „Information in Foto gefunden“, „Information in Text gefunden“ und „Information in Graphik/Tabelle gefunden“ (1 = L, 2 = LE, 3 = CL).

Tabelle 12-10 Mehrfachvergleich für "Element in dem Info gefunden wurde"

Mehrfachvergleiche							
Tamhane							
Abhängige Variable	(I) V3	(J) V3	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
gewünschte Information in Foto	1	2	2,0870 <sup>*</sup>	,2373	,000	1,508	2,666
		3	1,3774 <sup>*</sup>	,2075	,000	,872	1,883
	2	1	-2,0870 <sup>*</sup>	,2373	,000	-2,666	-1,508
		3	-,7096 <sup>*</sup>	,2771	,035	-1,382	-,037
	3	1	-1,3774 <sup>*</sup>	,2075	,000	-1,883	-,872
		2	,7096 <sup>*</sup>	,2771	,035	,037	1,382
gewünschte Information in Text	1	2	1,8788 <sup>*</sup>	,2575	,000	1,251	2,507
		3	1,1132 <sup>*</sup>	,2714	,000	,453	1,774
	2	1	-1,8788 <sup>*</sup>	,2575	,000	-2,507	-1,251
		3	-,7655 <sup>*</sup>	,1938	,000	-1,236	-,295
	3	1	-1,1132 <sup>*</sup>	,2714	,000	-1,774	-,453
		2	,7655 <sup>*</sup>	,1938	,000	,295	1,236
gewünschte Info in Grafik/Tabelle	1	2	-,0465	,1783	,991	-,480	,387
		3	,6226 <sup>*</sup>	,2248	,020	,077	1,168
	2	1	,0465	,1783	,991	-,387	,480
		3	,6691 <sup>*</sup>	,1947	,003	,195	1,144
	3	1	-,6226 <sup>*</sup>	,2248	,020	-1,168	-,077
		2	-,6691 <sup>*</sup>	,1947	,003	-1,144	-,195

\*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

Bei der Suchaufgabe vom Typ *look up* gab ein großer Teil der Personen (83,0 %) an, die gesuchte Information in einem Foto gefunden zu haben. Nach einer dahingehenden Analyse der Videos wurde deutlich, dass damit nach der in dieser Arbeit verwendeten Einordnung (siehe Klassifikation für Bilder, Anhang 1) eine Infografik gemeint war. Dabei handelte es sich allerdings nicht um die gesuchte *Dauer* der Fahrt, sondern um die zurückgelegte *Strecke*. Da aber vereinbart war, dass in die Erledigung der Aufgaben während des Experiments nicht eingegriffen werden darf, wurden die Nutzer nicht darauf hingewiesen. Für die Untersuchungen selbst spielt dies insofern keine Rolle, als dass einheitlich alle Benutzer das vermeintliche Suchergebnis in der Fahrstrecke sahen und der Aufgabentyp, der für die Untersuchungen die wesentliche Komponente darstellt sowie die Art der zu findenden Information dadurch nicht verändert wurden.

Auch beim Vergleich der Typen *learn* und *look up* zeigten sich statistisch signifikante Unterschiede für zwei der drei Inhaltselemente (Text und Foto) in denen die gewünschte Information gefunden wurde (siehe Tabelle 12-10)

Der Vergleich zwischen *learn* und *casual-leisure* ergibt statistisch signifikante Ergebnisse für zwei Elemente (Grafik und Text). Bei den *learn* Aufgaben waren für die Aufgabe relevante Informationen in allen Inhaltselementen enthalten. Eine genaue Interpretation der Daten ist allerdings insofern schwierig, als gerade bei den *casual-leisure* Aufgaben keine „gewünschte Information“ im einheitlichen Sinne formuliert war, sondern die Information „gewünscht“ war, die den Leser in diesem Moment interessiert. Es gibt demnach auch bei der freien Suche unterschiedliche Präferenzen.

Das bedeutet in der Konsequenz, dass diese Angaben bei *casual-leisure* den persönlichen Vorlieben entsprechen, da die gesuchten Informationen „freiwillig gewählt“ wurden. Bei den *learn* Aufgaben gab es ein bestimmtes Lernziel zu erreichen. Die dazu verwendeten Inhaltselemente oblagen aber der Präferenz des Nutzers und dessen subjektiver Relevanzbewertung. Folglich ist davon auszugehen, dass beim Typ *learn* die Wahl der Inhaltselemente vom Lernziel und der persönlichen Relevanz, bei den *casual-leisure* Aufgaben vor allem von den persönlichen Faktoren beeinflusst wurden.

### 12.3.5 Angaben zum Suchergebnis und Suchverlauf

Im Mittel wurde die Frage, ob die gewünschte Information gefunden wurde, mit nur 2,41 Punkten bewertet ( $n = 168$ ). Dieser eher niedrige Wert ist vermutlich auch der begrenzten Zeit geschuldet, die den Probanden zur Verfügung stand. Da die Suchtasks nicht als direkte Fragestellung formuliert (siehe dazu auch 11.2.3) und auch keine Ergebnisse bzw. ein Lernerfolg nach Ende des Experiments abgefragt wurden, entspricht die Einschätzung, ob die gewünschten Informationen gefunden wurden, der subjektiven Auffassung des Nutzers.

Auch die Zufriedenheit mit dem Suchverlauf wurde nur mit 2,35 Punkten im Mittel angegeben ( $n = 168$ ).

Aus den erhobenen Daten lässt sich keine Information darüber entnehmen, weshalb die Personen mit dem Suchverlauf nicht zufrieden oder zufrieden waren, denn weder der Wunsch nach mehr Bildern und/oder der Wunsch nach mehr Text und/oder nach Infografiken korreliert statistisch signifikant mit der Zufriedenheit des Nutzers (*look up*: Korrelationskoeffizienten Foto: 0,059 Text: 0,184, Infografik -0,347; *learn*: Korrelationskoeffizienten Foto -0,240, Text: -0,020, Infografik: -0,433, *casual-leisure*: Korrelationskoeffizienten Foto: 0,340, Text: -0,222, Infografik: -0,158).

Es wurde weiterhin untersucht, wie sich die Mittelwerte der Angaben für „mehr von einem Inhaltstyp gewünscht“ über die Gruppe der eher zufriedenen und der eher unzufriedenen VPs verhalten. Hier zeigte sich lediglich ein Zusammenhang bei dem Wunsch nach mehr Infografiken (t-Test, Gruppentrennwert 4 bei Variable Zufriedenheit,  $n_{\text{zufrieden}} = 110$ ,  $n_{\text{weniger zufrieden}} = 58$ , max. erlaubte Irrtumswahrscheinlichkeit 3,3%).

Tabelle 12-11 t-Test Zufriedenheit mit dem Suchverlauf und dem Wunsch nach einem bestimmten Inhaltstyp

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. [2-seitig]	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
mehr Text gewünscht	Varianzen sind gleich	3,582	,060	-,051	166	,960	-,0109	,2161	-,4376	,4157
	Varianzen sind nicht gleich			-,046	52,626	,964	-,0109	,2399	-,4922	,4704
mehr Fotos gewünscht	Varianzen sind gleich	,179	,672	1,982	166	,049	,4381	,2210	,0018	,8743
	Varianzen sind nicht gleich			1,995	60,863	,051	,4381	,2196	-,0011	,8772
mehr Graphiken/Tabellen gewünscht	Varianzen sind gleich	2,804	,096	5,070	166	,000	-1,0927	,2155	-1,5182	-,6672
	Varianzen sind nicht gleich			4,590	52,944	,000	-1,0927	,2380	-1,5702	-,6152

Es besteht ein genereller Zusammenhang zwischen der Zufriedenheit des Nutzers und damit, ob die Präsentation als angemessen bewertet wurde (Korrelationskoeffizient 0,782):

Tabelle 12-12 Korrelation zwischen der Zufriedenheit mit dem Suchverlauf und angemessener Präsentation

Korrelationen		
	Präsentation angemessen	Suchverlauf zufrieden
Korrelation nach Pearson	1	,782**
Präsentation angemessen Signifikanz [2-seitig]		,000
N	168	168
Korrelation nach Pearson	,782**	1
Suchverlauf zufrieden Signifikanz [2-seitig]	,000	
N	168	168

\*\* Die Korrelation ist auf dem Niveau von 0,01 [2-seitig] signifikant.

Dieses Ergebnis bestärkt die Hypothese dieser Arbeit, dass eine angemessene Präsentation der Inhalte einen starken Einfluss auf die Nutzerzufriedenheit hat.

Ebenfalls erkennbar ist ein Zusammenhang zwischen der Präsentation der Information und dem (subjektiven) Erfolg der Testperson:

*Tabelle 12-13 Korrelation zwischen dem Sucherfolg und der Bewertung, ob die Präsentation der Inhalte angemessen war*

Korrelationen		Präsentation angemessen	gewünschte Information gefunden
Präsentation angemessen	Korrelation nach Pearson	1	,800**
	Signifikanz [2-seitig]		,000
	N	168	168
gewünschte Information gefunden	Korrelation nach Pearson	,800**	1
	Signifikanz [2-seitig]	,000	
	N	168	168

\*\* . Die Korrelation ist auf dem Niveau von 0,01 [2-seitig] signifikant.

Der Korrelationskoeffizient von 0,8 zeigt auf dem Niveau von 0,01 signifikant einen starken (stärker als in der ersten Studie) Zusammenhang zwischen den beiden Einschätzungen der Nutzer. Auch wenn aus den Nutzerdaten der Fragebögen nicht weiter zu erschließen war, *welche* Inhalte jeweils konkret als angemessen erschienen, so ist doch eindeutig, dass der inhaltliche Erfolg der Suche stark von einer *angemessenen* Informationspräsentation abhängt. Diese Information ist wichtig für eine Argumentation pro adaptive Informationspräsentation.

## 12.4 Auswertung der Videolabels

Die annotierten Videos bzw. die exportierten Labels wurden vorverarbeitet wie in der ersten Studie: Die Annotationen wurden als .xml Datei exportiert und durch den *Logfile-Processor*<sup>62</sup> in 50ms Frames unterteilt. Es wurde wieder mit den an der Anzahl der Interaktionen normierten Häufigkeiten für die einzelnen Videolabel-Kombinationen gearbeitet, da auch in der zweiten Studie die Tasks in der Gesamtlänge variierten. Insgesamt konnten 101 Datensätze entsprechend 101 Aufgaben zur Analyse verwendet werden.

<sup>62</sup> Programm, das extra zur parallelen Verarbeitung der Anvil-Dateien, der FireFox-Logs und der Logfiles des Eyetrackers erstellt wurde, um allen 50ms Frames ein Event zuzuordnen.

Auf Basis der Ergebnisse der ersten Studie wird für die zweite Studie die Hypothese aufgestellt, dass sich die Informationsrezeption und die Nutzeraktionen abhängig vom Grad der Definiertheit und der Endlichkeit des Suchziels verändern. Dabei bezieht sich die Hypothese vor allem auf die Lektüre von Text, die Berücksichtigung von Bildern und des Inhaltsverzeichnisses.  $H_0$  ist also, dass es *keine* Mittelwertsunterschiede in der Population gibt, bzw. dass diese zufällig und nicht für die Grundgesamtheit zutreffend sind (alle Stichproben stammen aus der selben Grundgesamtheit).

In diesem Fall können gerichtete Hypothesen formuliert werden (basierend auf den Erkenntnissen der ersten Studie).

$H_1$  = Ist das inhaltliche Suchziel endlich und stark definiert, wird weniger Text gelesen.

$H_2$  = Ist das inhaltliche Suchziel endlich und stark definiert, werden weniger Bilder berücksichtigt.

$H_3$  = Ist das inhaltliche Suchziel endlich und stark definiert wird, das Inhaltsverzeichnis stärker berücksichtigt.

$H_4$  = Ist das inhaltliche Suchziel endlich und stark definiert wird, die Einleitung weniger berücksichtigt.

Es wurde ein Kruskal-Wallis-Test durchgeführt, um die Verteilung der Inhaltselemente zwischen den drei Klassen zu vergleichen. Es sind vergleichbare (bzw. die den deutschen Bezeichnungen entsprechenden Videolabel-Kombinationen) unterschiedlich verteilt wie bereits in der ersten Studie (RE-IN, RE-TX, RE-IX, SC-TX, EX-PI, SC-IB). Die Unterschiede betreffen vor allem Häufigkeit, wie oft Text gelesen wurde, wie viele Bilder im Fokus standen, wie häufig das Inhaltsverzeichnis und die Einleitung zur Problemlösung herangezogen wurden.

Tabelle 12-14 Kruskal-Wallis-Test über die Verteilung aller Videolabel-Kombinationen über alle drei TT

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von RE-LI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,275	Nullhypothese behalten.
2	Die Verteilung von RE-IN ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,001	Nullhypothese ablehnen.
3	Die Verteilung von RE-TX ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,005	Nullhypothese ablehnen.
4	Die Verteilung von RE-HL ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,092	Nullhypothese behalten.
5	Die Verteilung von RE-IX ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,017	Nullhypothese ablehnen.
6	Die Verteilung von RE-LINK ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,072	Nullhypothese behalten.
7	Die Verteilung von RE-IG ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,249	Nullhypothese behalten.
8	Die Verteilung von SC-LI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,268	Nullhypothese behalten.
9	Die Verteilung von SC-TX ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,012	Nullhypothese ablehnen.
10	Die Verteilung von SC-IN ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,066	Nullhypothese behalten.
11	Die Verteilung von SC-HL ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,589	Nullhypothese behalten.

Asymptotische Signifikanz werden angezeigt. Das Signifikanzniveau ist ,05.

	Nullhypothese	Test	Sig.	Entscheidung
12	Die Verteilung von SC-IX ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,226	Nullhypothese behalten.
13	Die Verteilung von SC-LINK ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,058	Nullhypothese behalten.
14	Die Verteilung von EX-PI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,003	Nullhypothese ablehnen.
15	Die Verteilung von SC-IB ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,033	Nullhypothese ablehnen.
16	Die Verteilung von EX-IB ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,500	Nullhypothese behalten.
17	Die Verteilung von EX-IG ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,092	Nullhypothese behalten.
18	Die Verteilung von NA ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,000	Nullhypothese ablehnen.
19	Die Verteilung von SC ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,813	Nullhypothese behalten.

Asymptotische Signifikanz werden angezeigt. Das Signifikanzniveau ist ,05.

Auch bei der Inspektion der Mittelwerte der relativen Häufigkeiten bestätigten die Ergebnisse die der ersten Studie. Es wurde über die unterschiedlich verteilten Variablen als post-Hoc-Test ein Mehrfachvergleich (Games-Howell-Test<sup>63</sup>) durchgeführt, wobei der Faktor der Tasktyp ist (max. erlaubte Irrtumswahrscheinlichkeit  $10/4 = 2,5$  also max. 2,5 % bzw. 0,025). Vergleichbar mit den Ergebnissen der ersten Studie zeigen sich hier Unterschiede bei den Mittelwerten von: Lesen der Einleitung (RE-IN), der Lektüre von Text (RE-TX), dem Lesen des Inhaltsverzeichnisses (RE-IX). Auch bei der Berücksichtigung von Bildern (EX-PI) finden sich statistisch signifikante Unterschiede.

<sup>63</sup> Da keine Varianzhomogenität zwischen den Gruppen vorliegt, siehe dazu Anhang 8.4.

Tabelle 12-15 Mehrfachvergleich (Games-Howell) der unterschiedlich verteilten Nutzeraktionen

Mehrfachvergleiche							
Games-Howell							
Abhängige Variable	(I) V1	(J) V1	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
RE-IN	1	2	-,07481 <sup>*</sup>	,01797	,001	-,1203	-,0293
		3	-,10241 <sup>*</sup>	,02370	,002	-,1649	-,0400
	2	1	,07481 <sup>*</sup>	,01797	,001	,0293	,1203
		3	-,02761	,02411	,503	-,0905	,0353
	3	1	,10241 <sup>*</sup>	,02370	,002	,0400	,1649
		2	,02761	,02411	,503	-,0353	,0905
RE-TX	1	2	-,26663 <sup>*</sup>	,07021	,006	-,4528	-,0805
		3	-,17846	,07054	,062	-,3658	,0089
	2	1	,26663 <sup>*</sup>	,07021	,006	,0805	,4528
		3	,08817	,04749	,179	-,0325	,2088
	3	1	,17846	,07054	,062	-,0089	,3658
		2	-,08817	,04749	,179	-,2088	,0325
RE-IX	1	2	,06258	,03503	,219	-,0322	,1574
		3	,09970 <sup>*</sup>	,03245	,036	,0074	,1920
	2	1	-,06258	,03503	,219	-,1574	,0322
		3	,03712 <sup>*</sup>	,01397	,048	,0003	,0739
	3	1	-,09970 <sup>*</sup>	,03245	,036	-,1920	-,0074
		2	-,03712 <sup>*</sup>	,01397	,048	-,0739	-,0003
EX-PI	1	2	-,07913 <sup>*</sup>	,01958	,003	-,1302	-,0281
		3	-,03304	,01590	,141	-,0761	,0101
	2	1	,07913 <sup>*</sup>	,01958	,003	,0281	,1302
		3	,04609	,02367	,152	-,0138	,1060
	3	1	,03304	,01590	,141	-,0101	,0761
		2	-,04609	,02367	,152	-,1060	,0138

Um die Unterschiede zwischen den einzelnen Aufgaben zu untersuchen wurde weiterhin jeweils die Verteilung der Mittelwerte der relativen Häufigkeiten zweier Tasktypen miteinander verglichen (siehe Tabelle 12-15). Das Lesen der Einleitung (RE-IN) und von Text (RE-TX) sowie ebenfalls das Betrachten von Bildern (EX-PI) ist statistisch signifikant unterschiedlich verteilt zwischen *look up* (1) und *learn* (2). Entsprechend  $H_1$  und  $H_2$  und  $H_4$  ist die Mittelwertdifferenz ( $\mu_1 - \mu_2$ ) von  $p(\text{RE-TX})$ : - 0,27;  $p(\text{RE-IN})$ : - 0,07 und  $p(\text{EX-PI})$ : - 0,08 negativ (siehe Tabelle 12-15). Also liegt bei (1) der Mittelwert aller Stichprobenmitglieder unter dem von (2) und entsprechend  $H_3$  ist die Mittelwertdifferenz zwischen der Suchaufgabe mit stark definiertem und endlichem Suchziel im Vergleich zu der Aufgabe mit weniger stark definiertem und nur sehr schwach endlichem Suchziel positiv: Mittelwertdifferenz  $p(\text{RE-IX})$ : 0,06, allerdings zwischen (1) und (2) nicht mehr mit einem akzeptablen alpha-Fehler-Niveau ( $p = 0,219$ ).

Zwischen *look up* (1) und *casual-leisure* (3) ergab sich (neben der bereits gezeigten unterschiedlichen Häufigkeit der Navigationszeiten) für das Lesen der Einleitung (RE-IN),



der Lektüre von Text (RE-TX) und des Inhaltsverzeichnisses (RE-IX) und der Berücksichtigung von Bildern (EX-PI) folgende Ergebnisse:

Nach  $H_1$  und  $H_2$  und  $H_4$  sollte auch hier wieder die Mittelwertdifferenz ( $\mu_1 - \mu_3$ ) von (1) und (3) negativ sein  $p(\text{RE-TX})$ : - 0,18;  $p(\text{EX-PI})$ : - 0,03 und  $p(\text{RE-IN})$ : - 0,10 und entsprechend  $H_3$  die Mittelwertdifferenz von  $p(\text{RE-IX})$ : 0,10, also positiv.

Die statistischen Auswertung zwischen den Tasktypen *learn* (2) und *casual-leisure* (3) zeigte bei den Mittelwerten der Wahrscheinlichkeit für das Auftreten der Videolabel-Kombinationen Unterschiede für die Lektüre des Inhaltsverzeichnisses ( $\mu_2 - \mu_3$ )  $p(\text{RE-IX})$ : 0,04. Da bei *learn* (2) zunächst von einem stärker definierten und endlicheren Suchziel ausgegangen wird als bei *casual leisure* (3) trifft auch hier  $H_3$  zu, dass bei höherem  $S_1$  und  $S_2$  das Inhaltsverzeichnis mehr berücksichtigt wird. Die andern Hypothesen ( $H_{1,2,4}$ ) treffen beim Vergleich zwischen (2) und (3) nicht zu. Der Grund dafür kann darin liegen, dass bei den *casual-leisure* Aufgaben der Nutzer ein persönliches inhaltliches Ziel formuliert haben kann, so dass sich – in diesem Fall –  $S_1$  und  $S_2$  der beiden Aufgabentypen annähern. Damit stammen die Stichproben dann tatsächlich teilweise aus der gleichen Grundgesamtheit.

Die auf Basis der ersten Studie aufgestellten Hypothesen konnten mit der zweiten Studie teilweise angenommen werden.

Einen genaueren Eindruck dieser Unterschiede vermitteln die folgenden Grafiken der zwischen den Aufgabentypen am stärksten nicht gleich verteilten Aktivitäten.

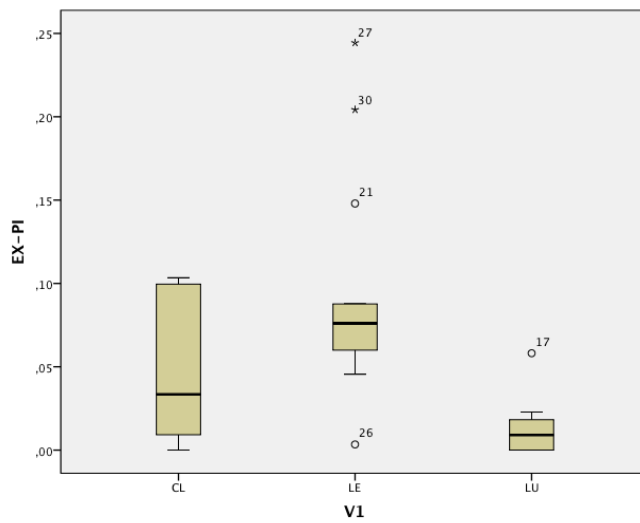


Abbildung 12-3 relative Häufigkeiten EX-PI Verteilung alle drei Tasktypen Boxplot

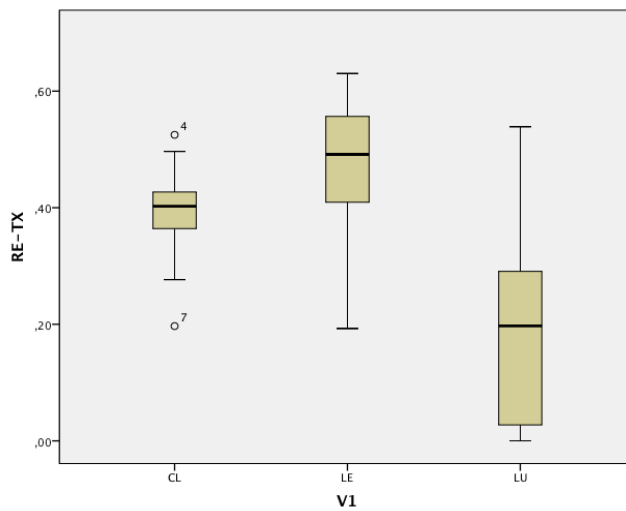


Abbildung 12-4 relative Häufigkeiten RE-TX Verteilung alle drei Tasktypen Boxplot,

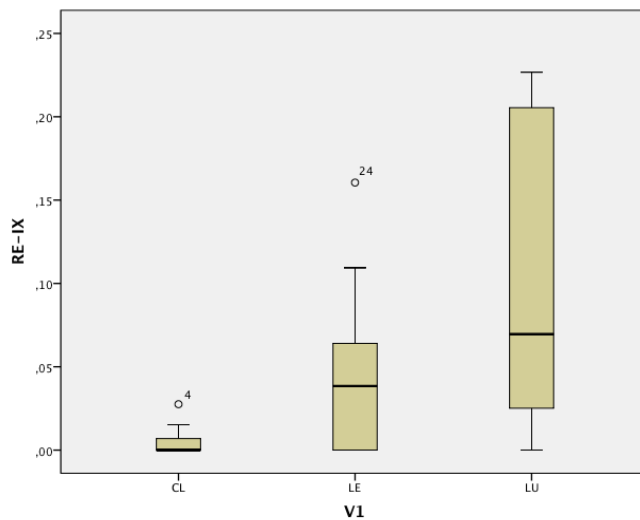


Abbildung 12-5 relative Häufigkeiten RE-IX Verteilung alle drei Tasktypen Boxplot

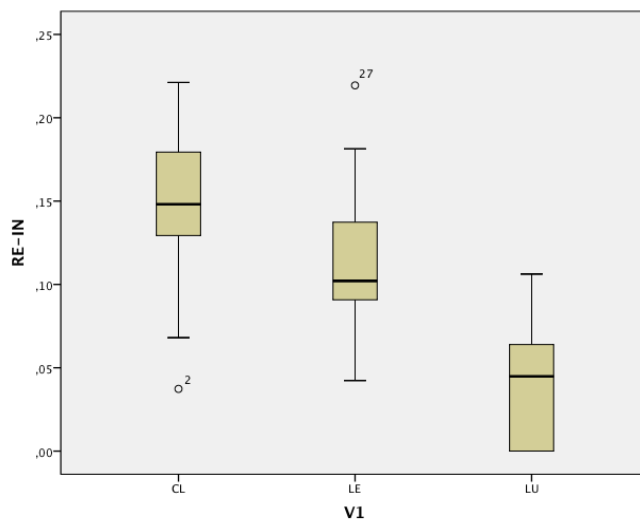


Abbildung 12-6 relative Häufigkeiten RE-IN Verteilung alle drei Tasktypen Boxplot

Die relativen Häufigkeiten, dass Bilder angesehen werden, sind bei allen drei Tasktypen verhältnismäßig<sup>64</sup> klein, bei den *casual-leisure* Aufgaben liegt die relative Häufigkeit in einem etwas breiteren Bereich, bei den *learn* Szenarien variieren die relativen Häufigkeiten nicht ganz so sehr, dafür zeigt das Diagramm (siehe Abbildung 12-3) Extremwerte. Bei *look up* werden Bilder grundsätzlich weniger berücksichtigt ( $h_m(\text{EX-PI}) = 0,00 - 0,02$ ). Die relativen Häufigkeiten für das Lesen von Text ist, wie der Boxplot (siehe Abbildung 12-4) zeigt, bis auf Ausreißer bei *casual leisure* eher homogen. Bei *learn*

<sup>64</sup> Das ist auch darauf zurück zu führen, dass das Betrachten eines Bildes im Gegenteil zur Lektüre von Text relativ wenig Zeit in Anspruch nimmt.

wird relativ unterschiedlich viel gelesen und bei *look up* schwankt die gelesene Textmenge sogar stark ( $h_n(\text{RE-TX}) = 0,00 - 0,54$ ).

Weiterhin zeigen die Grafiken (siehe Abbildung 12-5), dass bei den *casual-leisure* Aufgaben der Index der Artikel kaum Berücksichtigung fand. Beim explorativen Stöbern wurde hier offensichtlich direkt in den Inhalten gesucht. Bei den *look up* Szenarien dagegen ist die Verwendung des Indexes breiter gestreut ( $h_n(\text{RE-IX}) = 0,00 - 0,23$ ). Hier trat also auch eine verhältnismäßig starke Berücksichtigung des Inhaltsverzeichnisses auf. Das ist dadurch zu erklären, dass dort Hinweise darauf gesucht wurden, in welchem Artikelabschnitt vermutlich die gewünschte Information gefunden werden kann. Je schneller der entsprechende Teil des Artikels identifiziert werden konnte, desto kürzer war die Zeit, die mit Lektüre des Inhaltsverzeichnisses verbracht wurde. Bei den *learn* Szenarios liegen die relativen Häufigkeiten von (RE-IX) auch in einem niedrigen Bereich, das Inhaltsverzeichnis wurde aber mehr gelesen als bei den freien Aufgaben.

Abbildung 12-6 zeigt die Verteilung der relativen Häufigkeiten der Lektüre der Einleitung. Die Streuung ist zwischen den drei Aufgabentypen nicht so stark unterschiedlich, die Häufigkeiten sind allerdings bei CL am höchsten. Als Grund dafür ist zu vermuten, dass bei den freien Aufgaben die Nutzer ein Bild über die Inhalte und ob sie sich für den Artikelgegenstand interessieren anhand der Einleitung bildeten bzw. ohne Lernziel allgemeine und überblicksartige Informationen mehr akzeptabel waren als bei einem definierten Lernziel. Bei LU sind die relativen Häufigkeiten für eine Berücksichtigung der Einleitung am niedrigsten und auch generell niedrig. Hier waren allgemeine Informationen zum Artikelgegenstand offensichtlich weniger stark relevant bei der Informationsrezeption. Die relativen Häufigkeiten für (RE-IN) bei LE liegen zwischen den der beiden Aufgabentypen.

#### 12.4.1 Korrelationen Videolabels

Es wurde weiter überprüft, ob die Nutzeraktionen (in Form der Videolabels) miteinander korrelieren (also z.B. der Frage nachgegangen, ob die Lektüre von Text mit der Zeit der Bildbetrachtung, zusammenhängt). Es gibt hier zwar einige Effekte zu berichten, die wenigstens davon sind allerdings statistisch signifikant (siehe Anhang 8.10). In beiden Studien zeigte sich eine schwache signifikante Korrelation zwischen der Verwendung von Bildern und der Lektüre von Text (Studie eins Korrelationskoeffizient nach Pearson: 0,300, Studie zwei 0,372). Weitere schwache, statistisch signifikante Kor-

relationen zeigen sich generell zwischen Überblicks-schaffenden Aktionen wie dem Scannen verschiedener Inhaltselemente untereinander (siehe dazu Anhang 8.10).

#### 12.4.2 Zusammenfassung der Auswertung

Fasst man die Ergebnisse aus beiden Studien zusammen, dann bestätigt sich vor allem wieder die unterschiedliche Verwendung von Bildern und Text in den verschiedenen Suchaufgaben. Wie die vorangegangenen Grafiken und der Kruskal-Wallis-Test zeigen gibt es auch bei der Verwendung des Inhaltsverzeichnisses Unterschiede.

Es konnten auf Basis der ersten Studie gerichtete Hypothesen aufgestellt und durch die Daten der zweiten Studie teilweise bestätigt werden. Das bestärkt die generelle Annahme, dass die Informationsrezeption bei der Suche in der Wikipedia abhängig vom Nutzungsszenario, hier vor allem dem Grad der Definiertheit und der Endlichkeit des Suchziels variiert. Für eine adaptive Inhaltsauswahl ist die Kenntnis dieser Unterschiede eine wichtige Grundlage.

In der zweiten Studie fielen die Unterschiede zwischen den einzelnen Szenarien weniger stark aus als in der ersten Studie. Beim Vergleich der Mittelwerte der drei Gruppen gab es vor allem zwischen *learn* und *look up* Unterschiede zu beobachten.

In der zweiten Studie konnten vor allem auch zusätzlich wichtige Informationen über die Wünsche der Nutzer bezüglich bestimmter Inhaltstypen und über ihre subjektive Einschätzung der Suche und des Suchverlaufs sowie Zusammenhänge zwischen der Zufriedenheit der Nutzer und der Informationspräsentation bei der Suche in der Wikipedia gewonnen werden.

### 12.5 Weitere Arbeiten

Im Rahmen der Untersuchungen und Forschungen dieser Dissertation fanden und finden weitere Arbeiten statt. Einige davon stehen in direktem Zusammenhang mit dieser Arbeit, z.B. um weitere offene Punkte des Einflusses der Nutzerszenarien auf das Suchverhalten zu klären bzw. Erkenntnisse Szenario-spezifischer Interaktionen zu gewinnen. Alle Arbeiten wurden von der Autorin betreut und es wurden bis auf die

Untersuchungen in 12.6 jeweils die Daten aus der zweiten hier vorgestellten Studie verwendet.

### 12.5.1 Zusammenhang zwischen durchschnittlicher Sakkadenlänge und Task-Typ

Eine Hypothese, die mit den Studierenden [Adlfinger et al., 2013] aus dem Seminar im WS 2012/2013 an der Universität Regensburg aufgestellt wurde, war, dass die durchschnittliche Sakkadenlänge<sup>65</sup> bei den Suchaufgaben, in denen gezielt nach einer definierten Information gesucht wird, länger ist als bei Aufgaben, bei denen die Testperson gebeten wurde, sich so intensiv wie möglich über ein bestimmtes Thema zu informieren, also die Endlichkeit der Information [S<sub>1</sub>] und die Definiertheit [S<sub>2</sub>] relativ niedrig waren. Die Idee dahinter ist, dass durch das „Überfliegen“ der Inhalte auf den Wikipedia-Seiten längere Sprünge der Augen entstehen als z.B. beim Lesen oder wenn Bilder ausführlich betrachtet werden. Erkenntnisse in diesem Bereich könnten dabei hilfreich sein, anhand der Blickbewegungen des Nutzers Rückschlüsse auf sein Suchziel zu gewinnen.

Für die Auswertungen zu dieser Fragestellung wurden die Log-Dateien verwendet, die der Eyetracker während des Experiments aufzeichnet. Darin sind Informationen über die Blickbewegungen enthalten wie: Fixationen und Sakkaden (incl. Start- und Endposition (Koordinaten) sowie Anfangs- und Endzeit), Lidschläge und darüber hinaus auch verschiedene Nutzerevents (z.B. keypress). Aus diesen Daten berechneten die Seminarteilnehmer die Längen der Sakkaden mit Hilfe der entsprechenden Formel<sup>66</sup>:

$$\text{Sakkadenlänge} = \sqrt{(\text{Start Loc. X} - \text{End Loc. X})^2 + (\text{Start Loc. Y} - \text{End Loc. Y})^2}^{67}$$

Sie verglichen diese, nach einer Normierung der Daten, um den Einfluss von nutzerabhängigen Schwankungen zu vermeiden, hinsichtlich der verschiedenen Tasktypen.

---

<sup>65</sup> Die Blickbewegungen des Menschen bestehen in der Regel aus Sakkaden und Fixationen, wobei Sakkaden als sprunghafte Augenbewegungen beschrieben werden. Sie wechseln sich mit den sogenannten Fixationen ab, bei denen das menschliche Auge meist nahezu bewegungslos auf einen bestimmten Punkt fokussiert ist und so Informationen erfassen kann [Radach et al., 2004]. Im sakkadischen Zustand hingegen kann der Mensch nur wenig bis gar keine Information aufnehmen.

<sup>66</sup> Um den Betrag der Länge der Sakkaden zu erhalten, da die Differenz der Koordinaten-Werte auch negativ sein, falls der Augensprung von unten nach oben gerichtet war.

Es zeigte sich, dass zwar im Diagramm sichtbare Unterschiede in der Verteilung der Sakkadenlängen zwischen den unterschiedlichen Suchaufgaben im Bereich der längeren und längsten Sakkaden (nach der Normierung Länge zwischen 0 und 1, länger also im Sinne von ca. 0,7 – 1) gibt, aber ein Signifikanztest ergab, dass keine statistisch signifikanten Unterschiede der Verteilung der Sakkadenlängen zwischen den Tasktypen nachweisbar sind [Adlfinger et al., 2013].

### 12.5.2 Einflussfaktor thematisches Vorwissen und Erfahrung mit der Wikipedia

In einer weiteren Seminararbeit von Scherr et al. [Scherr et al., 2013] wurden die Daten entsprechend der Theorie [White et al., 2009; Wildemuth, 2004; Marchionini, 1989] nach einem Einfluss des Vorwissens auf die Suchstrategien der Probanden untersucht. Die Selbsteinschätzung der thematischen Vorkenntnis der Testpersonen erfolgte auf einer sechsstufigen Likertskala (*Pre-Task*-Fragebogen, siehe dazu Anhang 7.2) . Nach diesem Wert wurden die Nutzer in zwei Gruppen (0-1 = unerfahren, 2-5 = erfahren) eingeteilt. Zur Untersuchung des Interaktionsverhaltens wurden ebenfalls die Video-labels verwendet.

[Scherr et al., 2013] zeigen, dass statistisch signifikante Unterschiede zwischen den beiden Gruppen erkennbar waren, betreffend Anzahl/Minute der fixierten Überschriften, Inhaltsverzeichnisse, Einleitungsabsätze so wie die geklickten Hyperlinks, die zu einer nächsten Seite führten.

Mit 9,9% und 9,5% ist der wahrscheinlichste Übergang von TX (Text fixiert) zu ÜS (Überschrift fixiert) bei beiden Gruppen, gefolgt von ÜS (Überschrift fixiert) zu TX (Text fixiert) (9,0%) bei themenerfahrenen Personen und GR (Grafik fixiert) zu GR (Grafik fixiert) (7,5%) bei themenunerfahrenen Personen. Der Übergang mit der dritthöchsten Wahrscheinlichkeit ist bei den themenerfahrenen Personen ÜS (Überschrift fixiert) zu ÜS (7,5%) und bei den themenunerfahrenen Personen ÜS zu TX (6,9%). Demnach konzentrieren sich themenerfahrene Personen eher auf Überschriften und Texte, wohingegen themenunerfahrene Personen häufiger zusätzlich Grafiken betrachten [Scherr et al., 2013].

Der interessanteste Aspekt bei den Untersuchungen von [Scherr et al., 2013] ist, dass der Wechsel der Zustände (welches Inhaltselement wird fixiert) von Themen-unerfahrenen Personen auf einer Seite statistisch signifikant höher ist als bei Themen-erfahrenen Nutzern. Diese Ergebnisse zeigen, dass sich Themen-unerfahrene Nutzer in bestimmten Punkten anders verhalten als solche, die bereits Vorkenntnisse haben.

### 12.5.3 Gelesene Textmenge in Abhängigkeit vom Tasktyp

Einer ähnlichen Fragestellung wie in den Auswertung der Studien in dieser Arbeit gingen auch Forster et al. 2013 [Forster et al., 2013], ebenfalls im genannten Seminar, nach:

Sie analysierten die gelesene Textmenge der einzelnen Tasks anhand der mittels des Eyetrackers erstellten Videos. Gezählt wurde dabei nur das Lesen von Fließtext, nicht wenn der Nutzer Infografiken, Tabellen oder Infoboxen, die mitunter ebenfalls Text enthalten, betrachteten. Auf Basis von [McNamara, 2001] gingen die Autoren davon aus, dass die Lesezeit mit der Textmenge korreliert. Es wurde somit auf Basis der festgehaltenen jeweiligen Lesezeit während der Tasks auf die Textmenge rückgeschlossen. Es zeigten sich unterschiedliche Mittelwerte für die relativ zur Taskdauer verbrachte Lesezeit, vor allem für den Tasktyp *look up*, wo im Mittel nur 23,56 % der Zeit damit verbracht wurden, Fließtext zu lesen. Beim Tasktyp *learn* wurde im Mittel am meisten gelesen, nämlich 52,17 % der Zeit (*casual leisure*: 49,80 % der Zeit). Auch hier zeigt sich der Einfluss von  $S_1$  (Definiertheit des Suchziels) auf das Verhalten der Nutzer, vor allem auf die gelesene Textmenge.

## 12.6 Einfluss der Domain auf die Präferenzen für Inhaltselemente

Die Inhalte in der Wikipedia sind breit gefächert (siehe dazu auch 9.7.5). Darstellung und Erklärung von Programmiersprachen, Sehenswürdigkeiten und Artikel zu Personen sind nur ein kleiner Teil davon. Nach [Yan et al., 2011; Yue et al., 2013; Calderon 2012 und Han et al., 2013] kann auch die durchsuchte Wissensdomain die Präferenzen und das Suchverhalten der Nutzer beeinflussen.

In den im Rahmen dieser Arbeit durchgeführten Studien eins und zwei konnten keine umfassenden Unterschiede in der Präferenz für bestimmte Inhaltselemente zwischen



den gleichen Tasktypen mit unterschiedlichen Themen festgestellt werden. Lediglich bei *look up* aus Studie zwei unterschied sich die Häufigkeit, mit der eine Infografik betrachtet wurde, statistisch signifikant zwischen den beiden thematisch unterschiedlichen *look up* Aufgaben. Das zeigte bereits die Auswertung der Fragebögen, in denen die Frage danach, in welchem Inhaltstyp die gewünschte Information gefunden wurde, sehr häufig mit „Bild“ beantwortet wurde (womit aber nach Auswertung der Videos offensichtlich eine Infografik gemeint war). Somit ist der Zusammenhang darin zu sehen, dass die gesuchte Information dieser Infografik (siehe Abbildung 12-7, es war nach der *Fahrzeit* der Bahn gesucht. Die Probanden missinterpretierten allerdings die Streckenlänge als gesuchte Information, diese war *auch* im Text zu finden. Dennoch wurde hier die Infografik stark bevorzugt) zu entnehmen war und weniger die direkte Abhängigkeit von der Thematik.

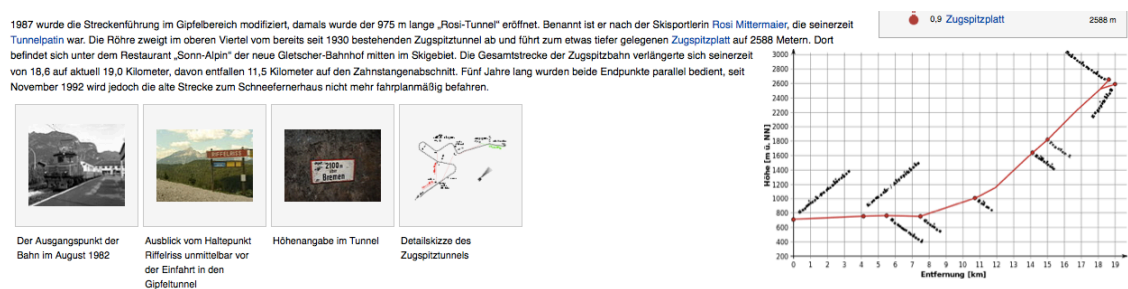


Abbildung 12-7 Ausschnitt aus dem Artikel über die Bayerische Zugspitzbahn in der deutschsprachigen Wikipedia [abgerufen: 12.Juli 2013, Quelle: [http://de.wikipedia.org/wiki/Bayerische\\_Zugspitzbahn](http://de.wikipedia.org/wiki/Bayerische_Zugspitzbahn)]

In einer weiteren Studie im Rahmen einer Masterarbeit [Vogel, 2013]<sup>68</sup> wurde deshalb der Einfluss der Domain auf die Präferenzen für die Inhaltselemente in der deutschsprachigen Wikipedia untersucht.

Beim Design der Studie zur Untersuchung der Präferenzen für bestimmte Inhaltstypen in Abhängigkeit der durchsuchten Domain wurden mit Hilfe des hier erstellten Kontextmodells aus Kapitel 6.2 ausschließlich Tasks formuliert, die in dem Typ *learn* ent-

<sup>68</sup> Unter Betreuung der Autorin.

sprechen (Also:  $M = 4-5$ ,  $S_1 = 1-2$ ,  $S_2 = 2-3$  und  $S_3 = 0-1^{69}$ ). Daraus ergaben sich Aufgaben, die keine bestimmten Elemente bevorzugten (z.B. weil nur nach einer Zahl, einem Text oder einem Bild gesucht werden konnte/sollte), aber auch keine Elemente ausschlossen (die Versuchsperson sollte sich einen Überblick über das Thema schaffen, einen gewissen Spielraum in der Interpretation boten die Formulierungen der Tasks nach [Borlund, 2000]).

### 12.6.1 Aufbau der Studie

Für die Untersuchungen wurde wiederum der Eytracker verwendet (siehe Kapitel 11.1.5.2 für Details). Allerdings kamen diesmal statische Seiten zum Einsatz, also die entsprechenden Artikel aus der Wikipedia im PDF-Format. Das bot den Vorteil, dass sog. *Areas of Interest* (AoI) im Vorfeld der Untersuchungen definiert werden konnten. Für die Benennung der AoIs wurden entsprechend die im Rahmen dieser Arbeit definierten Labels der Videoannotation (siehe Kapitel 11.3.2) verwendet, um Vergleichbarkeit mit den bisherigen Ergebnissen zu schaffen.

Die Testpersonen wurden gebeten, sechs Suchaufgaben zu bearbeiten. Die Themen der Aufgaben wurden nach der Klassifikation (siehe Kapitel 9.7.4) ausgewählt und zugeordnet.

Die verwendeten Artikel betrafen: *die Terroranschläge des 11. Septembers*, die Sängerin *Amy Winehouse*, die Autorin *Astrid Lindgren*, die Sportart *Bogenschießen*, das *Sonnensystem* und die Insel *Bali*. Die Themen wurden jeweils mit der Klassifikation aus Kapitel 9.7.4 bestimmten Domains zugeordnet.

Wikipedia-Artikel	Domain	Artikelgegenstand
Bali	Geographie	Insel
Astrid Lindgren	Literatur	Person
Bogenschießen	Sport	Methode
Sonnensystem	Naturwissenschaft	Ort
Terroranschläge vom 11.September 2001 <sup>45</sup>	Politik	Ereignis
Amy Winehouse	Musik	Person

<sup>69</sup> Auch hier hängen die genauen Ausprägungen der Variablen von der Interpretation der Suchaufgaben durch die Testperson ab.

Abbildung 12-8 Zuordnung der Untersuchungsgegenstände zur Klassifikation, [Vogel, 2013, s.34]

Der Bekanntheitsgrad der Themen wurde vorher abgefragt, um auch hier einen Einfluss (durch stark vorhandenes oder gänzlich fehlendes Vorwissen) zu berücksichtigen.

Bei der Auswertung der Daten zeigte sich, dass die Angaben zum Vorwissen ausreichend homogen waren, um einen Einfluss dieses Faktors auszuschließen [Vogel 2013].

### 12.6.2 Ergebnisse

Insgesamt wurden 16 Personen getestet. Die Teilnehmer der Studie waren zwischen 20 und 35 Jahre alt, sechs davon waren männlich, zehn weiblich. Alle Probanden hatten entweder bereits einen Hochschulabschluss oder befanden sich noch im Studium. Es zeigte sich auch in dieser Studie, dass Erfahrung im Umgang mit und Bekanntheitsgrad der Wikipedia hoch waren [Vogel, 2013].

Hier seien exemplarisch die Auswertungen der Artikel Bali (Insel:Geografie) und Amy Winehouse (Person:Musik) aufgeführt. Für eine vollständige Auswertung siehe [Vogel, 2013]. Gezählt wurden die Millisekunden und die Häufigkeiten, wie oft eine AoI angesehen wurde. Da für die Lösung der Aufgaben immer gleichviel Zeit zur Verfügung stand, mussten die Werte nicht an der Zeit normiert werden (hier: Summe<sup>70</sup> *Duration* Thema A: 2169883 ms, Summe *Duration* Thema B: 2194837ms).

---

<sup>70</sup> Summe der ms aller Probanden für die jeweiligen Artikel.

Die Anzahl der Wechsel zwischen den AoIs ist als absolutes Vorkommen von Interesse, auch hier ist die Summe der Häufigkeiten ähnlich und entsprechend der Zeit: Thema A: 6657 Wechsel (Summe aller VPs), Thema B: 6358 Wechsel (Summe aller VPs). Der Chi-Quadrat-Verteilungs-Test zeigt eine signifikant unterschiedliche Verteilung der Betrachtungsdauer der einzelnen Inhaltstypen.

*Tabelle 12-16 Verteilung der Betrachtungsdauer auf die AoIs Thema A und Thema B und Chi-Quadrat-Verteilungstest über Summe und Dauer*

AOI	DurationA	SummeA	DurationB	SummeB
HL	111961	381	73539	243
IB	0,00001	0,00001	100505	267
IG	39182	123	180186	495
IN	143061	466	37508	100
LI	68959	195	166556	452
PI	61505	201	75941	225
TX	1745215	5291	1560602	4576
	Chi-Quadrat über Duration	Chi-Quadrat über Summe		
p=	0,00E+00	0,00E+00		

Die nachfolgenden Grafiken zeigen die unterschiedliche Verteilung der Verwendung der einzelnen Inhaltselemente.

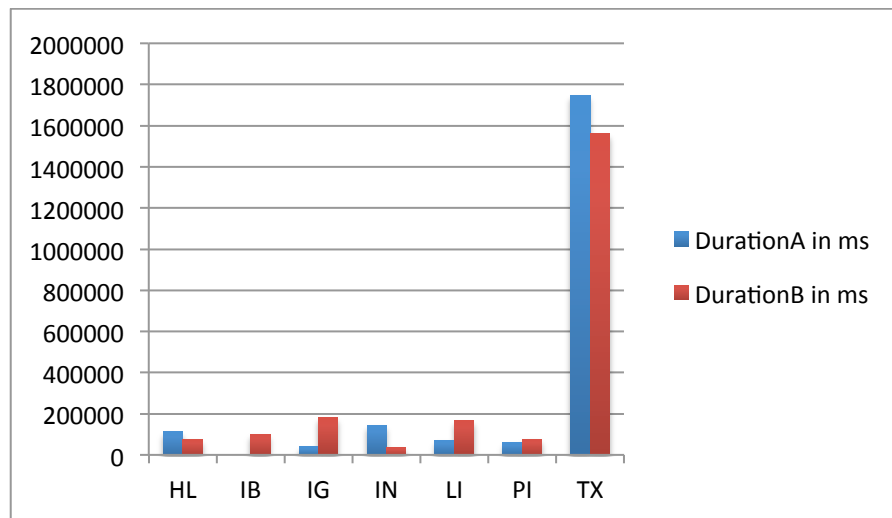


Abbildung 12-9 Dauer der Betrachtung der Inhaltselemente Thema A und Thema B Summe aller VPs

Die Grafik ohne das Inhaltselement Text (TX) verdeutlicht die Unterschiede besser, da die durch die hohe Summe der Dauer der gelesenen Textmenge ausgelöste Stauchung wegfällt. Grafische Elemente wurden bei Aufgabe Bali stärker berücksichtigt als bei Aufgabe Amy Whinehouse. Hier wurde der Artikel-Text so wie die Einleitung und die Überschriften mehr gelesen, also generell alle textuellen Elemente außer der Liste. Bei der Suche nach Informationen über die Insel Bali schienen den Probanden die Informationen aus Infografiken und Listen offensichtlich hilfreicher, während bei der Sängerin der Text die gewünschten Informationen zu bieten hatte.

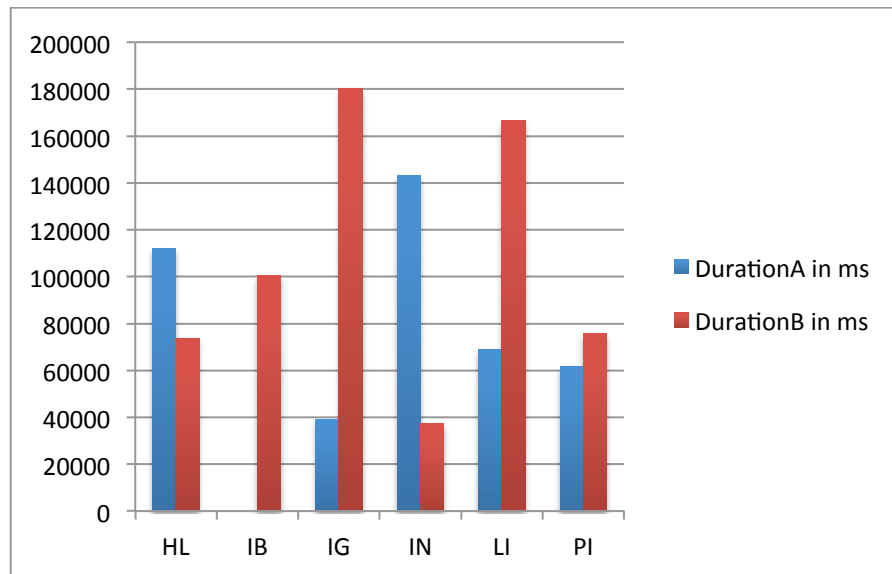


Abbildung 12-10 Dauer der Betrachtung der Inhaltelemente Thema A und Thema B Summe aller VPs ohne TX

Die Ergebnisse von [Vogel, 2013] bestätigen auch die unterschiedliche Verwendung von Bildern. Interessant dabei ist, dass vor allem bei den beiden Artikeln über Personen die Summe über alle Probanden, in denen Bilder betrachtet wurden relativ gering ist. Bei der Aufgabe zum Thema *Sonnensystem* wurde deutlich mehr Zeit als bei allen anderen Aufgaben damit verbracht, Bilder zu betrachten.

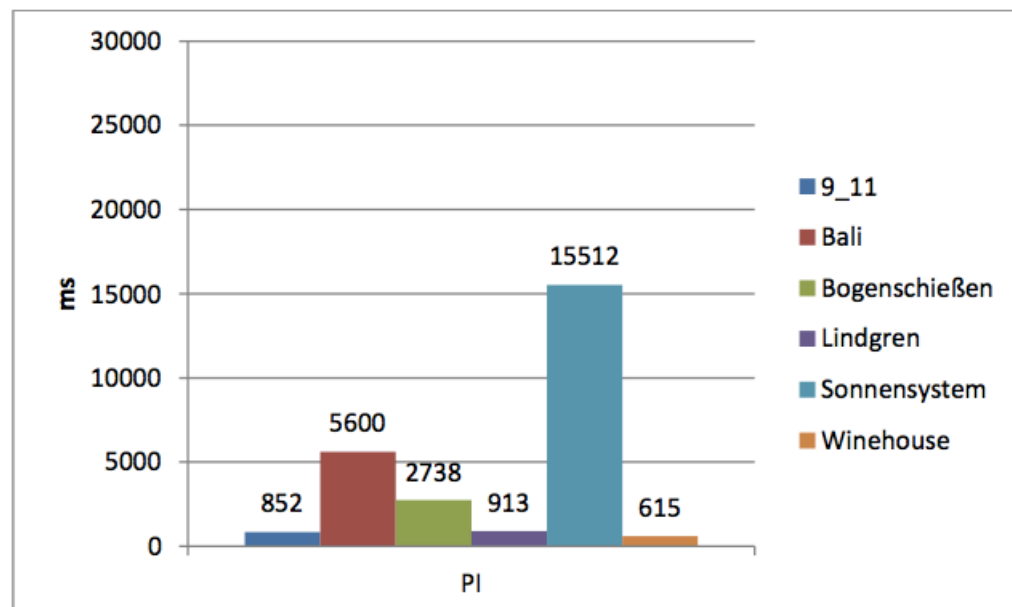


Abbildung 12-11 Fixationsdauer des Inhaltelements Bild in Summe aller VPs [Vogel, 2013, S. 51]

In der vollständigen Auswertung der Daten bei [Vogel, 2013] zeigte sich, dass bei der Suche in der Wikipedia Unterschiede im Suchverhalten und den Präferenzen für In-

haltselemente abhängig von der durchsuchten Domain nachzuweisen sind. Dabei ist zu berücksichtigen, dass die Unterschiede vermutlich nicht ausschließlich durch die jeweilige Thematik der Artikel hervorgerufen wurde, aber da die Aufgabentypen konstant gehalten wurden, Einflussfaktoren wie Vorwissen, Interesse und Selbstauskunft zur emotionalen Lage relativ homogen waren, so wie darauf geachtet wurde, dass das Verhältnis der Inhaltselemente der verschiedenen Artikeln entsprechend war, kann von einem Einfluss des Themas auf die Informationsrezeption ausgegangen werden, wenngleich auch nicht so stark, wie die hier dargestellten Daten zunächst vermuten lassen.





### 13 Auswertung des Modells

Die Erkenntnisse aus den Studien werden auf das Modell angewendet. Im Rahmen dieser Arbeit nicht nachweisbare Einflussfaktoren bleiben weiterhin Teil des Modells, da diese, wie der Forschungsliteratur zu entnehmen ist, in anderen Zusammenhängen bereits überprüft worden sind (siehe dazu Kapitel 5.6). Die Auswertung bezieht sich hier konkret auf die im Rahmen dieser Arbeit durchgeführten Studien.

Tabelle 13-1 Einfluss der Variablen des Modells<sup>71</sup>

#	Name	Einflussbereich	Einfluss
V <sub>1</sub>	Motivation	Navigationszeit, Berücksichtigung des Inhaltsverzeichnisses, Wunsch nach bestimmten Inhaltselementen	schwach bis mittelstark
V <sub>2</sub>	Suchziel (S <sub>1</sub> und S <sub>2</sub> )	Lektüre von Text, Berücksichtigung von Bildern, Lektüre der Einleitung, Berücksichtigung des Inhaltsverzeichnisses und der Einleitung, Wunsch nach bestimmten Inhaltselementen	stark
V <sub>2</sub>	S <sub>3</sub> nicht informationsbezogenes Suchziel	Navigationszeit, Berücksichtigung des Inhaltsverzeichnisses, Wunsch nach best. Inhaltselementen	schwach bis mittelstark
V <sub>3</sub>	Stimmung	Wunsch nach Bildern	mittelstark
V <sub>4</sub>	Zufriedenheit	korreliert mit Informationspräsentation	stark
V <sub>5</sub>	Sucherfolg	abhängige Variable	
V <sub>6</sub>	Vorkenntnis V <sub>Thema</sub>	Berücksichtigung des Indexes	schwach
V <sub>6</sub>	Vorkenntnis V <sub>WP-Recherche</sub>	Aktionswechsel	schwach
V <sub>x</sub>	persönliche Präferenzen	hier nicht erfasst	
V <sub>D</sub>	V <sub>D1</sub> = Alter	homogene Altersgruppe	
V <sub>D</sub>	V <sub>D2</sub> = Geschlecht	kein Einfluss nachweisbar	

<sup>71</sup> Legende: Grün = starke Zusammenhänge erkennbar, gelb = schwache Zusammenhänge erkennbar

#	Name	Einflussbereich	Einfluss
V <sub>C</sub>	Komplexität	homogene Komplexität	
V <sub>7</sub>	Domain	Präferenzen für Inhaltselemente	mittelstark
V <sub>8</sub>	verwendetes Gerät	bei allen Aufgaben gleich	
V <sub>9</sub>	verfügbare Zeit	bei allen VPs gleich	
V <sub>10</sub>	Aktion	abhängige Variable	
V <sub>11</sub>	task completion	abhängige Variable	
V <sub>12</sub>	betrachtetes Element	abhängige Variable	
V <sub>12B</sub>	Art des Bildes	Berücksichtigung und Bewertung von Bildern	schwach bis mittelstark
V <sub>13</sub>	Relevanz (E)	abhängige Variable	
V <sub>14</sub>	Suchphase	Berücksichtigung der Inhaltselemente	mittelstark

Die Auswertung zeigt, dass in den hier durchgeführten Studien teilweise Einflüsse der Variablen des Modells auf die Informationsrezeption und das Nutzerverhalten bei der Suche in der Wikipedia nachgewiesen werden konnten. Die direkten Zusammenhänge bzw. Auswirkungen sind den jeweiligen Auswertungen der Studien zu entnehmen. Um differenziertere Eindrücke über die einzelnen Einflüsse zu erhalten, sollten diese zukünftig jeweils einzeln (z.B. Einfluss des Suchthemas etc.) weiter untersucht werden. In den hier durchgeführten Studien war vor allem das Ziel, explorativ vorzugehen und die Frage, ob generell konkrete Einflüsse des Modells auf das Nutzerverhalten bei der Informationssuche in der Wikipedia nachweisbar sind, zu untersuchen.

## 14 Diskussion der Ergebnisse

In dieser Arbeit wurde der Frage nach heterogenen Szenarien bei der Nutzung der Wikipedia und der generellen Überlegung, ob eine Berücksichtigung dieser bei der Informationspräsentation sinnvoll wäre, nachgegangen.

Die Auswertung der Forschungsliteratur dazu zeigte zunächst auf breiter Basis die generelle Existenz unterschiedlicher Szenarien beim Information Retrieval (Siehe ab Kapitel 5).

Die Frage, ob auch die Wikipedia Gegenstand unterschiedlicher Nutzungssituationen ist, konnte auf Basis der Literatur und durch empirische Untersuchungen (siehe Kapitel 9) bejaht werden.

Deshalb wurde im nächsten Schritt der Frage nachgegangen, ob eine Berücksichtigung unterschiedlicher Nutzungsszenarien bei der Wikipedia sinnvoll sein könnte. Zur Klärung dieser Frage wurden verschiedene Aspekte untersucht. Dazu zeigten erstens empirische Studien am Beispiel des Medientyps Bild (siehe Kapitel 10), dass Inhalte nicht nur nach ihrer semantischen Richtigkeit sondern nach weiteren Kriterien wie persönlichen Präferenzen und Funktion ausgewählt werden. Diese subjektive Bewertung von Inhalten spricht für eine situativ adaptierte Inhaltspräsentation.

Ebenso unterstützt zweitens die Theorie des *cognitive viewpoint* [Ingwersen und Järvelin, 2005] und die der situativen Relevanz [u.a. Buckland, 1991; Kuhlthau, 1991; Ingwersen 1982] die Berücksichtigung des Nutzers und seines Kontextes für ein effektiveres Information Retrieval.

Es wurde drittens gezeigt, dass durch Adaption bereits Erfolge bei der Verbesserung der IR-Ergebnisse erzielt wurden (siehe Kapitel 7.2.2) und Adaption auf Basis der Nutzerinteraktion generell ein probates Mittel ist, um die Nutzerzufriedenheit zu verbessern (siehe Kapitel 7.2)

Die offene Frage nach situativen Präferenzen der Nutzer für bestimmte Inhaltselemente bei der Suche in der deutschsprachigen Wikipedia wurde empirisch untersucht.

Anhand zweier Studien wurden teilweise Unterschiede in der Informationsrezeption bei verschiedenen Suchaufgaben-Typen nachgewiesen. Auch wenn nicht alle erfassten Faktoren aus dem Kontext des Nutzers (siehe dazu Kapitel 13) einen hier nachweisba-

ren Einfluss auf das Verhalten bei der Suche zeigten, so waren doch Zusammenhänge erkennbar (siehe dazu ab Kapitel 11).

Damit kann die Hypothese, dass das Suchverhalten der Nutzer abhängig vom Nutzungsszenario variiert (bzw. die dazugehörigen Einzelhypothesen), angenommen werden. Die einzelnen Fragestellungen und Untersuchungen werden im Folgenden kurz besprochen.

### ***Verschiedene Nutzungsszenarien und deren Berücksichtigung***

Die Untersuchungen dieser Arbeit wurden motiviert durch die Vermutung, dass es unterscheidbare, beschreibbare Nutzungsszenarien gibt und abhängig von diesen die Vorgehensweise und die Wünsche der Nutzer sich beim Information Retrieval ändern. Es wurde vor allem deutlich gemacht, dass die Frage nach dem Nutzungskontext mehrdimensional ist.

Es spielen unterschiedliche Faktoren in der Situation des Nutzers, in der er einen Information Retrieval Prozess startet und durchführt, eine Rolle. Das Zusammenspiel dieser Faktoren und die Vorstellungen des Nutzers über seine Problemlösungsstrategien fügen eine persönliche Komponente zu den Standardverfahren des Information Retrievals hinzu (siehe auch [Belkin, 1993]). Diese persönliche Komponente einer Informationsrecherche in digitalen Informationssystemen steuert das Suchverhalten, die Informationsrezeption und Vorstellung des Nutzers davon, wie er die Suche kompletieren oder seinen Informationsbedarf stillen will [Wang und Soergel, 1998].

Die Situation des Information Retrievals ist aus Büros oder Bibliotheken zusätzlich in weitere Umgebungen gelangt. Die Orte, Umgebung und Intention der Suchenden sind also ausgeweitet und damit, in Kombination mit dem „Problem“ der enorm großen Datenmengen, die weltweit zur Verfügung stehen, ist der situative Aspekt der Suche und die Selektion von Information wichtiger geworden.

Mit Unterstützung der Theorien und Modelle des *Information Seeking* und *Information Behaviour* wurde veranschaulicht, dass sowohl die situativen Aspekte aus dem Kontext als auch die Interaktion mit den jeweiligen Inhalten den Nutzer weiter beeinflussen. Dieser Einfluss ist vor allem in einer jeweils subjektiven und situativen Relevanzbe-

wertung und Informationsrezeption der Inhalte zu sehen. Die Evaluierung der z.B. von einer Suchmaschine zurückgelieferten Ergebnisse erfolgt nach bestimmten Kriterien, die Subjektivität findet also vor allem an *dieser* Stelle des *Information Seeking* Prozesses statt.

Konkrete Faktoren, die die Bewertungen der Inhalte beeinflussen wurden ab Kapitel 5.4 gesammelt und besprochen. Es wurde gezeigt, dass formal unterscheidbare Suchziele vorhanden sind, auch wenn diese nicht immer in ihrer Reinform auftreten. Deshalb wurden Klassen von Suchzielen verwendet. Die Grundlage der situationsrelativen Bewertungen der Ergebnisse liegt also im auslösenden Task. Der Grund, der Auslöser und die Motivation für eine Suche determinieren damit neben der kognitiven Struktur Ziele und Erwartungen des Nutzers. Aus der Literatur geht hervor, dass unterschiedliche Suchstrategien auf verschiedene Auslöser-Situationen folgen.

Gleiches gilt für die Motivation der Nutzer, die ab Kapitel 5.5 besprochen wurde. Der Fokus lag hier vor allem darauf, dass die Motivation dafür, zum einen generell online zu gehen und/oder zum anderen speziell nach Informationen zu recherchieren, nicht immer ausschließlich oder überhaupt durch einen konkreten Informationsbedarf entsteht. Verschiedene Untersuchungen wurden inspiziert, aus denen hervorgeht, dass auch andere Gründe wie die Überbrückung eines Zeitraums, Ablenkung, Entspannung oder Langeweile als Motivation für eine Suche existieren können. Es wurde damit dargestellt, dass es unterschiedliche Motivations-Typen für eine Suche gibt. Ist das Ziel aber nicht eine bestimmte Information, sondern zum Beispiel das Bedürfnis, die eigene Stimmung zu verbessern, Zeit zu vertreiben oder einfach nur Spaß zu haben, dann rücken andere Kriterien in den Vordergrund als bei der Erfüllung eines bestimmten Informationsbedarfs. Damit ist auch die affektive Ebene des Nutzers bei der Suche eine Zielgröße und der Zusammenhang zwischen den Inhalten und den Emotionen eines Nutzers spielt eine Rolle. Diese Erkenntnisse bestärken die Hypothese der situationsabhängigen Informationsrezeption.

Ebenso relevant sind die Erkenntnisse über die Rolle der kognitiven Struktur des Nutzers, wozu neben seinem (thematischen) Vorwissen auch seine Vorstellung von der Welt und den Dingen darin eine Rolle spielt. Nach dem *cognitive viewpoint* [Ingwersen

und Järvelin, 2005] ist es plausibel, dass diese vorhandene kognitive Struktur, Weltbild und Erwartungen der Nutzer einen wesentlichen Einfluss darauf haben, wie er ein IR-Problem zu lösen gedenkt und damit auch auf seine Informationsrezeption bzw. die subjektive Selektion von Inhalten. Dabei geht es verstärkt darum, um noch einmal [Belkin, 1980, S. 134] aufzugreifen:

*„to help people to solve problems, rather than directly to solve problems posed to them“*

Um aber das leisten zu können werden konkrete Erkenntnisse über die Erwartungen und Zusammenhänge von *task* und (inhaltlichen) Zielen sowie über die Rolle der weiteren Einflussfaktoren benötigt. Deshalb wurden hier die Auswirkungen und Einflüsse zunächst generell im IR auf Basis der Literatur und später empirisch für den Spezialfall der deutschsprachigen Wikipedia untersucht.

Eine entscheidende Erkenntnis ist dabei, dass, neben der Motivation, auch die *Ziele* der Nutzer anhand verschiedener Variablen voneinander abgrenzbar sind (siehe dazu Kapitel 5.4). Untersuchungen aus der Forschungsliteratur zeigten, dass eine Berücksichtigung dieser bei der Auswahl der Treffermenge die Bewertungen durch die Nutzer verbesserten (siehe dazu Kapitel 7.2). Auch dies bestätigt die Annahme der situationsrelativen Informationsrezeption.

Die Zusammenführung der einzelnen, identifizierten Einflussfaktoren auf die Informationsrezeption und die subjektive Relevanzbewertung erfolgte mit Hilfe des in Kapitel 6.1 erstellten Modells. Hiermit wurde weiterhin im Sinne einer Konzeptspezifikation vor allem auch die Vielfältigkeit und die Mehrdimensionalität der möglichen Nutzungsszenarien deutlich. Die Überführung der latenten Variablen des Nutzerkontexts bzw. der Nutzungsszenarien in ein Set aus messbaren Indikatoren unterstützte dann im weiteren Verlauf der Arbeit die empirischen Untersuchungen zu den Nutzungsszenarien in der Wikipedia. Bei diesen Untersuchungen wurden die vorliegenden Erkenntnisse aus der Literatur, die in das Modell eingeflossen sind, am „Spezialfall“ der deutschsprachigen Wikipedia untersucht. Nach Abschluss der Studien wurden die Ergebnisse im Modell ergänzt.

Vor der empirischen Untersuchung musste allerdings noch geklärt werden, ob eine Berücksichtigung der besprochenen Einflüsse aus der jeweiligen Situation des Nutzers sinnvoll ist im Sinne von höherer Nutzerzufriedenheit bzw. situativ besser bewerteten Ergebnissen.

Im konkreten Bezug auf die Wikipedia waren dabei drei wesentliche Aspekte zu berücksichtigen: Erstens ob Adaption generell im Information Retrieval zu einer Verbesserung der Ergebnisse und der Nutzerzufriedenheit führt, zweitens ob die Wikipedia Gegenstand der heterogenen Nutzungsszenarien ist und drittens, ob es überhaupt Nutzungsszenario-abhängige Unterschiede bei den Präferenzen gibt.

Die erste Frage konnte gut auf Basis der vorhandenen Forschungsliteratur beantwortet werden: Auch im Information Retrieval konnten verbesserte Ergebnis-Sets erzielt werden, wenn einzelne Faktoren aus dem Kontext des Nutzers berücksichtigt wurden. Es stellte sich hier allerdings heraus, dass vor allem semantische Erweiterungen getestet wurden (siehe Kapitel 7.2.2).

Dass aber die Interaktion des Nutzers mit Inhalten und System wertvolle Informationen über seine Absichten liefert [Adafre und Rijke, 2006], seine Zufriedenheit [z.B. Feild et al., 2010] sowie über seine subjektive Bewertung der Inhalte [z.B. Claypool et al., 2001] ist eine wichtige Grundlage für die Überlegungen zu einer adaptiven Informationspräsentation. Denn werden diese Informationen berücksichtigt und fließen in den Informations-Auswahl-Prozess des Systems mit ein, so wäre eine Verbesserung (auf Basis der oben genannten Erkenntnisse) realistisch vorstellbar. Neben Erkenntnissen über Zusammenhänge zwischen den *tasks* und den Erwartungen der Individuen ist das ein entscheidender Aspekt. Es wurde in dieser Arbeit dargelegt, dass diese Möglichkeiten bestehen und diese im Rahmen als implizites Feedback zu einer Verbesserung der Informationspräsentation herangezogen werden können.

Weiterhin wurde hier dargelegt, dass die situativen Aspekte von Relevanz und Information ebenfalls ein Grund für eine kontext- und nutzerrelative Inhaltsauswahl beim Information Retrieval sind. Eine allgemeine semantische Relevanz bzw. eine potentielle Informativität der Inhalte kann zwar objektiv mehr oder weniger einheitlich bestimmt werden [Buckland, 1991], aber die von demjenigen, der die Information für ein konkretes Ziel benötigt, vergebene Relevanz entsteht subjektiv aus der Situation her-

aus. Sie kann also nicht zwingend mit der allgemeinen, objektiv ermittelten Relevanz als identisch betrachtet werden. Abschätzungen über die subjektive Relevanz bestimmter Inhaltstypen bei der Suche in der Wikipedia waren das Ziel der anschließend durchgeführten empirischen Untersuchungen.

Die Frage, ob es plausibel sei, bei der Nutzung der Wikipedia ebenfalls von heterogenen Situationen, im Sinne der bisher identifizierten Faktoren die ein Nutzungsszenario darstellen, auszugehen, ist die zweite Annahme, die bestätigt werden musste, um den Sinn einer adaptiven Informationsauswahl bei der Wikipedia-Nutzung zu untermauern. Zunächst ist festzuhalten, dass die kognitive Struktur, die affektiven Faktoren und das Vorwissen *immer* die Erwartungshaltung eines Menschen prägen. Es ist also davon auszugehen, dass diese Einflussgrößen demnach auch bei der Wikipedia-Nutzung gelten.

Inhalte und Aufbau der Wikipedia bestärken, wie in dieser Arbeit in Kapitel 9.2 dargelegt wurde, darüber hinaus die Vermutung, dass auch hier zu Freizeit Zwecken gesurft und gesucht wird. Die Angaben der Versuchspersonen in den im Rahmen dieser Arbeit durchgeführten Studien bestätigten diese Annahme ebenfalls (Unterscheidung der Suchszenarien nach der Motivation). Dass die Wikipedia weiterhin sowohl für das Fakten-Retrieval als auch für ausführlichere Recherchen verwendet wird, zeigten [Lim, 2009 und Hu et al., 2009] (Unterscheidung der Suchszenarien nach dem Suchziel).

Zusammenfassend wurde damit gezeigt, dass die heterogenen Nutzungsszenarien auch auf die Wikipedia zutreffen. Auf dieser Basis wurde die Frage nach einer situationsrelativen Informationsrezeption in den genannten unterschiedlichen Situationen bearbeitet.

Die an dieses Thema anschließenden Fragen konnten nicht mehr auf Basis der Literatur beantwortet werden und wurden deshalb empirisch untersucht.

Bevor die generelle situationsabhängige Informationsrezeption bzw. Präferenz für bestimmte Inhalte bei der Suche in der Wikipedia untersucht wurde, bestand noch die Frage, welche Rolle die situative und subjektive Bewertung von Bildern spielt. Diese Frage leitete sich vor allem aus allgemeinen Erkenntnissen der Bildwissenschaft ab



(verschiedene Funktionen von Bildern, der Aspekt des „Gefallens“, siehe ab Kapitel 10). Deshalb wurden hier vor der Untersuchung der allgemeinen Informationsrezeption zunächst Bilder und deren persönliche Auswahl und Bewertung durch den Nutzer untersucht.

Zu dieser Frage wurden im Rahmen der Arbeit mehrere Untersuchungen vorgenommen (siehe dazu auch: [Knäusl et al., 2012b; Knäusl und Schubart, 2013; Schubart, 2013]). Die Ergebnisse dieser Forschungen und die Auswertung der Literatur zum Thema Bild (siehe u.a. [Nöth, 2008; Haasebrook, 1995; Schnotz, 2006; Peez, 2006 und Weidenmann, 1991]) zeigten, dass der Nutzer gerade bei der Bildauswahl stark von kognitiven, aber auch affektiven (siehe dazu auch [Schubart, 2013]) Faktoren geleitet wird. Es konnte gezeigt werden, dass bei Auswahl und Bewertung von Bildern durch den Nutzer nicht *nur* semantische Korrektheit entscheidend ist. Auch die Funktion eines Bildes, zum Beispiel die Frage, ob es ergänzend oder dekorativ verwendet wird, spielt eine Rolle bei der Entscheidung der Nutzer für oder gegen ein Bild. Ebenso hat die Tatsache, ob ein Bild gefällt, einen Einfluss auf die Auswahl, allerdings nicht ausschließlich, wie sich in Kapitel 10.8 zeigte. Bei den anschließenden Datenerhebungen zeigte sich auch weiterhin eine vom Aufgabentyp abhängige Berücksichtigung von Bildern bei der Problemlösung.

Dieser Aspekt unterstreicht das Potential einer situationsbezogenen Bildauswahl im Information Retrieval. Es wurde weiter besprochen, dass zu einer adaptiven Bildauswahl bestimmte Informationen über die Bilder vorliegen müssen, um eine geeignete Auswahl treffen zu können. In der Literatur finden sich verschiedene Versuche, Bilder automatisch auszuwählen, verschiedene Verfahren kommen dabei zur Anwendung (siehe dazu Kapitel 10.6). Faktoren wie die Emotionalität, der Bekanntheitsgrad eines Bildes oder das Ziel einer Suchaufgabe können dabei nicht berücksichtigt werden, aber aus der Literatur geht eindeutig hervor, dass Bilder nach verschiedenen, über die mitgelieferten Metadaten hinausreichende, Kriterien ausgewählt werden können.

Ob und wenn ja wie die ab Kapitel 5 genannten Faktoren eines Nutzungsszenario allgemein das Verhalten bei der Suche in der Wikipedia und der Auswahl der bei der Problemlösung berücksichtigten Inhalte beeinflussen, wurde ab Kapitel 11 untersucht.

In den beiden Studien wurde gezeigt, dass das Suchverhalten und die Informationsrezeption der Nutzer vor allem abhängig von Umfang und Definiertheit des inhaltlichen Ziels unterschiedlich ausfällt. Auch das Thema der Verwendung von Bildern tauchte hier wieder auf. Diese wurden in den Szenarien, in denen sog. Fakten-Retrieval betrieben wurde, wenig berücksichtigt, wogegen sie bei den ausführlichen Informationsrecherchen eine Rolle spielten.

Bei den nicht nur das Bestehen eines Informationsbedarf motivierten Aufgaben zeigte sich, dass die Menschen sich relativ unterschiedlich verhalten und es lange Phasen der Orientierung und des „Stöberns“ gibt. Weiter wurde deutlich, dass das Verhalten der Nutzer und ihre Interaktion mit den Inhalten nicht über die ganze Suche konstant sind. Auch hier fiel auf, dass Bilder, selbst wenn sie Berücksichtigung über die gesamte Suche fanden, zu Beginn der Suche eine untergeordnete Rolle spielten. Die situative Verwendung und Bewertung von Bildern bestätigte sich in verschiedenen Messungen wiederholt. Sowohl bei Befragung z.B. mittels Fragebogen (siehe Kapitel 12.3.3), als auch durch Beobachtung (mittels Videos, siehe Kapitel 11.3.4, 11.3.6, 11.3.7 und 12.4) waren Auftreten und Verwendung dieses Medientyps unterschiedlich.

Die automatische Zuordnung der Datensätze der Tasks (die die Interaktion mit den Inhalten betrafen) zeigte teilweise gute Ergebnisse. Für eine adaptive Informationspräsentation ist das eine notwendige Grundlage, da das System der Interaktion Informationen über die Absichten des Nutzers entnehmen können muss.

In der zweiten Studie konnten auf Basis der Erkenntnisse der vorangegangenen Untersuchungen gerichtete Hypothesen über die Verwendung bestimmter Inhaltselemente in den Suchszenarien aufgestellt werden. Die – unabhängig von der ersten Studie durchgeführte – zweite Erhebung mit anderen Aufgaben und neuen Versuchspersonen konnte einen großen Teil dieser Hypothesen bestätigen. Es manifestiert sich damit die Hypothese, dass die unterschiedliche Informationsrezeption nicht zufällig ist und auch auf die entsprechende Gesamtpopulation zutrifft.

Eine wichtige und die Motivation dieser Arbeit bestärkende Erkenntnis ist auch der starke Zusammenhang zwischen der Zufriedenheit der Personen mit der Informationspräsentation und ihrer generellen Zufriedenheit. Es bestätigt sich in den im Rah-

men dieser Arbeit durchgeführten Experimenten die Vermutung, dass nicht nur der letztendliche (inhaltliche) Erfolg, sondern auch der Verlauf der Suche und die Adäquatheit der präsentierten Information wichtig sind für die Zufriedenheit der Nutzer.

Gerade die eingangs genannten Faktoren wie ein zu hoher Aufwand oder eine zu hohe Anforderung an die Informationskompetenz der Nutzer, um die gesuchte und gewünschte Information zu finden, sind im Umkehrschluss zu den hier gezeigten Erkenntnissen ein Grund, dass die Nutzer weniger zufrieden sind.

Die Untersuchungen zu den weiteren genannten möglichen Einflussfaktoren zeigten nur teilweise Zusammenhänge. Die schnellere Abfolge an Zustandswechsel in der Interaktion von Personen mit weniger Themenerfahrung ist hier von Interesse, genauso wie die Ergebnisse einer weiteren Studie [Vogel, 2013], die deutlich machten, dass sich bei der Suche in der Wikipedia die Berücksichtigung der verschiedenen Inhaltstypen auch nach der Domain richtet.

Für den Einfluss der Emotionen konnten im Rahmen dieser Arbeit keine eindeutigen Beweise gefunden werden. Die emotionale Lage der Probanden änderte sich über den Zeitraum des Experimentes nur schwach, und auch zwischen dem Verhalten der Personen, die guter oder schlechter Grundstimmung zugeordnet wurden, zeigten sich keine signifikanten Unterschiede. Der Grund dafür ist einerseits bei der Methodik zu sehen - im Laborexperiment zur Wikipedia-Suche ließen sich keine starken affektiven Reaktionen oder Gefühlsschwankungen erzeugen. Die Erwartung, dass die Personen bereits zu Beginn des Experiments ausreichend unterschiedliche emotionale Ausgangssituationen mitbringen, wurde nicht erfüllt. Andererseits sind die auf Grund der Literatur vermuteten Einflüsse von Emotionen und Stimmung beim Information Retrieval – anders als bei z.B. Computerspielen – eher als Langzeitphänomen zu sehen [Nahl und Bilal, 2007].

Die Ergebnisse und Erkenntnisse dieser Arbeit zusammenfassend ist festzuhalten: Die Recherche in digitalen Informationssystemen ist Teil unseres Alltags geworden. Durch das Überangebot an Informationen und Inhalten ist die Suche trotz guter technischer Infrastruktur schwierig geworden. Die Tatsache, dass Informationsrecherche aus hete-

rogenen Gründen in den verschiedensten Kontexten betrieben wird, findet meist keine Berücksichtigung.

Auch die Suche in der Wikipedia ist Gegenstand unterschiedlich motivierter Suchszenarien, die wiederum jeweils unterschiedliche Ziele verfolgen. Dabei sind die Wünsche und Präferenzen der Nutzer in Abhängigkeit ihres Suchszenarios unterschiedlich. Gerade bei Bildern spielt die subjektive Bewertung des Nutzers und die Funktion, die das Bild im Text hat, eine Rolle dabei, ob es bei der Suche berücksichtigt wird oder nicht. Da Adaption und Berücksichtigung der Nutzersituation und der daraus entstehenden Bedürfnissen der Nutzer, die über die Erfüllung eines korrekten semantischen Abgleichs ihrer verwendeten Suchterme mit der Dokumentenmenge hinaus gehen, eine Verbesserung der Ergebnisse zeigen, ist diese Argumentation auch für die Verwendung der Wikipedia gültig.

## 15 Limitierungen der Arbeit

Diese Arbeit ist vor allem auf die Informationsrecherche in der Wikipedia konzentriert. Es wurde zusätzlich auf der Basis der allgemeinen Literatur gearbeitet. Die empirisch erhobenen Daten wurden hinsichtlich der jeweiligen Fragestellungen interpretiert.

Bestimmte Komponenten der Nutzungsszenarien zeigten in den vorgenommen Untersuchungen keine signifikanten Einflüsse auf das Nutzerverhalten. Das bedeutet, dass diese Einflüsse hier statistisch nicht nachweisbar waren, obwohl sie sich in anderen Untersuchungen (z.B. bei der Websuche) in der Forschungsliteratur zeigten.

Für machen Einflüsse, wie die Stimmung oder der Stresslevel sind dementsprechend Feldstudien den Laborstudien vorzuziehen. Für Einflüsse des thematischen Vorwissens und der Erfahrung der Nutzer mit dem verwendeten System bieten sich Langzeitstudien an wie von [Wildemuth, 2004], bei denen von den gleichen Wissensänderungen bei verschiedenen Personen auszugehen ist.

Das persönlich unterschiedliche Interaktionsverhalten zeigte sich bei allen im Rahmen dieser Arbeit durchgeführten Studien (in dieser Arbeit und bei [Vogel, 2013]). Auch wenn für einzelnen Interaktionsformen (wie z.B. mit der Maus, siehe dazu [Navalpakam und Churchill, 2012]) in der Literatur bereits Nutzertypen identifiziert wurden, so war es in diesen Studien nicht möglich, diese auch bei der Interaktion mit den Inhalten im Sinne der Videolabels zu identifizieren.

Bei der Auswertung der Logfiles, die die Interaktion der Nutzer mit dem Browser via Maus und Tastatur enthielten, konnten ebenfalls keine relevanten Zusammenhänge mit der jeweiligen Nutzeraktion (hinsichtlich der Inhalte) des Nutzers gefunden werden.

Die Annotation der Videos selbst war eine relativ subjektive Vorgehensweise (Beurteilung der jeweiligen Nutzeraktionen durch den Annotator). Die Verwendung statischer Seiten, bei denen mit Hilfe der Auswertungssoftware Areas of Interest und Blickpfade festgestellt werden können oder Kenntnis der genauen Koordinaten der einzelnen Inhaltslemente (im HTML-Code), kann hier methodische Verbesserung erwirken.

Bei der Aufgabenstellung für die Nutzer traten Probleme bei der Simulation der sog. *casual-leisure* Szenarien [Elsweiler et al., 2011] auf. Gerade in der zweiten Studie war die Aufgabenstellung nicht vollständig gelungen und löste bei den Nutzern nicht immer das intendierte, vom persönlichen Interesse geleitete, explorative Stöbern aus. Bei der Aufgabenstellung war der Gedanke, dem Nutzer in der Laborsituation die Chance zu geben, sich kurzzeitig mit persönlich interessanten Themen auseinander zu setzen als Alternative zu den simulierten Suchaufgaben initiativ ausschlaggebend, da auch die Literatur zeigt, dass ein Zusammenhang zwischen der Stimmung der Probanden und der Frage, ob sie einen selbstgewählten oder einen konstruierten Task bearbeiten müssen, vorhanden ist [Poddar und Ruthven, 2010]. Durch die hier gewählte Aufgabenstellung gelang das nur teilweise. In der ersten Durchführung der Datenerhebung zeigten die Nutzer exploratives Vorgehen nach persönlichem Interesse. Hier wurden sie aufgefordert, ein Thema auszuwählen, das sie interessiert. Bei der zweiten Studie wurde die Aufgabenstellung geändert, um noch stärker exploratives Verhalten zu provozieren, deshalb wurde die Eingabe eines Suchterms nicht mehr gestattet.

Die Erhebung der Daten in den empirischen Untersuchungen unterliegen Limitierungen.

Die Fragebögen wurden von den Nutzern selbst ausgefüllt, somit konnte nicht sichergestellt werden, dass alle Fragen gleichermaßen verstanden und beantwortet wurden.

Bei der Durchführung der Experimente konnten nur die jeweils zu den Themen vorhandenen Inhaltselemente betrachtet werden. Es wurde bei der Auswahl der Tasks darauf geachtet, dass Qualität und Quantität der Artikel entsprechend waren, aber da die Nutzer nicht auf einen bestimmten Artikel eingeschränkt wurden, navigierten sie auch über andere Seiten. Dort konnte das Angebot dann stark unterschiedlich sein.

In den Fragebögen wurden die Personen gebeten, anzugeben, ob sie sich jeweils mehr der einzelnen Inhaltstypen gewünscht hätten. Diese Angaben fielen alle bei allen Suchaufgaben relativ hoch aus (im Sinne von: „Ja, von allem mehr gewünscht“). Das kann unter anderem aber auch darauf zurück zu führen sein, dass die Artikel in der Wikipedia generell inhaltlich limitiert sind (z.B. im Gegensatz zu einer Websuche).

Die Daten der zwei Studien zum Suchverhalten in der Wikipedia wurden im Labor erhoben. Die Einschränkung einer Laborstudie gegenüber einer Feldstudie betrifft vor allem die Authentizität der Aufgaben, auch wenn durch die Formulierungen nach [Borlund, 2000] versucht wurde, größtmögliche Authentizität herzustellen. Der Nutzer fühlt sich stärker beobachtet als „im Feld“, wird u.U. von einer (fiktiven) Erwartungshaltung des Versuchsleiters beeinflusst, traut sich bei Missverständnissen nicht nachzufragen und/oder verhält sich eventuell nicht so, wie er es in einer natürlichen Situation tun würde. Andererseits bietet eine Laborstudie im Vergleich zu einer Feldstudie die bessere Möglichkeit, kontrollierte Daten aus mehreren Quellen zu erfassen. Hier wurde die Laborstudie vor allem auf Grund der Kontrollierbarkeit und dem Bedarf nach Archivierung des Blickverhaltens in Videoform (durch Nutzung eines Eyetrackers) durchgeführt.

Bei den im Rahmen dieser Arbeit in verschiedenen empirischen Studien erhobenen Daten handelt es sich jeweils um nicht repräsentative Stichproben der Gesamtpopulation. Weder in Alter noch Bildungsabschluss repräsentierten die Testpersonen die Bevölkerung. Der erwartete implizite Fehler ist damit hoch. Die Ergebnisse sollten nur unter Berücksichtigung dieser Einschränkung interpretiert werden.

Die Themengebiete der Aufgaben wurden so ausgewählt, dass es sich nicht um Spezialthemen handelte, bei denen erst Vorwissen notwendig gewesen wäre, noch um Themen, die keinerlei Interesse mehr wecken könnten, auszuschließen ist das aber nicht. Generell werden die Ergebnisse valider und informativer, je größer die Bandbreite der Themen und der Situationen und je heterogener die Nutzergruppe ist.





## 16 Fazit

Schon 1994 stellten Morita und Shinoda [Morita und Shinoda, 1994] einen Ansatz zur Berücksichtigung von persönlichen Faktoren beim Information Retrieval vor, der Verbesserungen bei der Ergebnisauswahl brachte. Motivation für die Autoren war, den Nutzer vor einem *information overload* [Morita und Shinoda, 1994] zu schützen.

Die Forschung dieser Arbeit und die besprochene Literatur zeigen, dass selektives Vorgehen der Nutzer in Abhängigkeit verschiedener Faktoren vorhanden ist. Problematisch dabei ist, dass die einzelnen Einflüsse oft latent, nicht messbar oder schwierig zu erfassen sind. Die Überführung dieser Einflüsse in ein konkretes Modell und das genau abgegrenzte Anwendungsgebiet der deutschsprachigen Wikipedia halfen, abgrenzbare Unterschiede und Einflüsse dennoch teilweise zu erfassen (siehe dazu auch Kapitel 11). Allerdings musste dabei ungeklärt bleiben, welchen Einfluss die Variablen untereinander haben, da hier die generelle Ableitung der konkreten Variablen aus der Literatur und eine Anwendung auf einen definierten Nutzungskontext im Vordergrund stand. Durch die Berücksichtigung der Literatur zum *Information Seeking Behaviour* generell und zum Einfluss von situativen Faktoren beim Information Retrieval wurde der Zusammenhang zur existierenden Forschung hergestellt bzw. auf vorhandenen Erkenntnissen aufgebaut, um ein realistisches Experimentdesign zu erstellen. Die Anwendung der gesammelten Erkenntnisse auf das Gebiet der Wikipedia-Suche schuf neue Erkenntnisse darüber, welche Faktoren relevant sind und welche konkreten Auswirkungen sie auf die Informationsrezeption der Nutzer haben. Diese konnten dann auf das eingangs auf Basis der bereits vorhandenen Theorien und Erkenntnisse erstellte Modell rückgeführt werden.

Die Studien zeigen, dass die Zufriedenheit der Nutzer nicht ausschließlich von der Erfüllung ihres Informationsbedarfs abhängt, sondern auch vom Suchverlauf und der Präsentation der Informationen. Eine Berücksichtigung der persönlichen Wünsche und Präferenzen hat folglich Potential zur Steigerung der Nutzerzufriedenheit.

Dazu sind aber immer genaue Kenntnisse über das jeweilige Informationssystem, dessen Zweck und die Inhalte notwendig (siehe dazu im Ausblick ab Kapitel 17 und im folgenden Kapitel).

Ein generelles Problem bei der Umsetzung einer adaptiven Informationspräsentation der Inhalte der Wikipedia ist allerdings der Datenschutz. Zu den Erkenntnissen in der Forschungsliteratur und denen dieser Arbeit konnte nur gelangt werden, da viele persönliche Daten über die Nutzer erhoben wurden. Die Sensibilität dieser Daten schwankt zwar, aber generell sind umfangreiche Informationen z.B. über die emotionale Lage des Nutzers oder seinen Aufenthaltsort keine Daten, die allgemein freizügig an Systeme übermittelt werden sollten.

Details wie das Frustrationsniveau oder der Verlauf der emotionalen Lage über die Suchaufgabe hinweg können u.U. sogar aus der Interaktion mit dem System erkannt werden (siehe dazu [Feild et al., 2010]). Um – vor allem privaten und in den sog. *casual-leisure* Szenarien (siehe dazu [ElsweilerWilson2010]), bei denen vor allem nicht-informationsbezogene Suchziele im Vordergrund stehen – optimale Adaptivität erreichen zu können, müssten diese Aspekte aber vermutlich bekannt sein.

Weitere Daten, wie der Einfluss der durchsuchten Domain, generelle Vorlieben (Lerntyp etc.) und der Einfluss des Tasktyps und des verwendeten Geräts (vor allem z.B. die Unterscheidung zw. PC und Smartphone) sowie der verfügbaren Zeit (hier vor allem vorstellbar bei Systemen im Alltag wie Ernährungsberater mit Kochrezepten, Stadtführer auf dem Smartphone oder Anwendungen aus dem eHealth Bereich etc.) könnten für private Zwecke möglicherweise der optimalen Informationspräsentation, gesteigerter Nutzerzufriedenheit und verbesserter Informationsrezeption zuliebe verwendet werden. Die Erstellung eines Nutzerprofils zur Unterstützung adaptiver Verfahren in einem geschützten Bereich eines Informationssystems ist nicht undenkbar.

Generell sollte die Verwendung privater Daten dringend am *Nutzen* der jeweiligen personalisierten und adaptiven Informationspräsentation gemessen werden.

Eine Inhaltsgenerierung zur Laufzeit hat zwar das Potential, Systeme nutzerfreundlicher zu gestalten, ob eine Umsetzung sinnvoll ist, bleibt vom betreffenden System ab-

hängig. Bei der Informationsrecherche in der Wikipedia wurde unterschiedliches Verhalten in Abhängigkeit der Ziele und Situation der Nutzer gezeigt und auf Basis der Literatur argumentiert, dass eine Berücksichtigung dieser Unterschiede generell sinnvoll ist. Da hier vor allem das inhaltliche Suchziel ausschlaggebend war für Unterschiede in den Präferenzen und der Informationsrezeption der Nutzer, sind nicht zwingend sensible Daten notwendig, um die Informationspräsentation zu optimieren. Wie die Experimente zeigten, können anhand der Nutzerinteraktion mit den Inhalten ab einem bestimmten Zeitpunkt der Suche Rückschlüsse auf einige Aspekte des Suchziels (z.B. sind umfangreiche Recherchen oder Fakten-Retrieval Ziel des Nutzers) gezogen werden. Eine Anpassung der präsentierten Inhalte auf Basis dieser Informationen benötigt keine sensiblen Daten über die reine Interaktion des Nutzers mit den Inhalten hinaus. Durch die vielfältigen Nutzungskontexte und Suchziele, die bei der Verwendung der Wikipedia gegeben sind, kann allein die Berücksichtigung des inhaltlichen Suchziels (und/oder der damit verbundenen Motivation zur Suche) durchaus bereits eine Verbesserung der Informationspräsentation erwirken, berücksichtigt man dazu auch, die Angaben, die die Probanden zu ihren *Wünschen* nach bestimmten Inhaltstypen gemacht haben. Diese unterschieden sich teilweise stark bei den Aufgabentypen.

Die entsprechende Auswahl von Bildern (in Abhängigkeit der bildinternen Faktoren über rein semantische Aspekte hinaus) ist dagegen tendenziell schwieriger umzusetzen. Hier spielt auch der persönliche Aspekt wie Bekanntheitsgrad des Bildinhaltes und der individueller Geschmack des Nutzers eine Rolle.

Zusammenfassend ist festzuhalten, dass die Frage, welche Aspekte in einer Anwendung konkret berücksichtigt werden sollten und wie die dafür notwendigen Informationen gewonnen werden, vom Nutzen und den Zielen eines Informationssystems abhängen. Die Motivation für eine angepasste und damit verbesserte Informationspräsentation bzw. Untersuchungen in diesem Zusammenhang ist vor allem die Zufriedenheit der Nutzer (durch besseres, schnelleres und passenderes Auffinden von Informationen).

---

Holistisch gesehen besteht aber die Zufriedenheit der Nutzer nicht nur aus deren Zufriedenheit mit der Informationspräsentation, sondern es müssen auch die informationelle Selbstbestimmung und der Datenschutz berücksichtigt werden.

## 17 Ausblick

Im Folgenden werden offene, an die Ergebnisse und Erkenntnisse der vorliegenden Arbeit anschließende Fragestellungen aufgeworfen und kurz erörtert.

### 17.1 Berechnung der Übergangswahrscheinlichkeiten

Bisher wurden bei der Analyse der Daten nur die Häufigkeiten bzw. Wahrscheinlichkeiten des Auftretens einer Videolabel-Kombination über einen kompletten Task berücksichtigt. Eine offene Frage ist, wie sich die Übergangswahrscheinlichkeiten zwischen den Zuständen (Nutzeraktionen) verhalten.

In [Knäusl und Ludwig, 2013] wurde versucht, die Übergangswahrscheinlichkeiten zwischen den Nutzeraktionen (entspricht den Videolabels) zwischen den 50ms Zeitspannen zu berechnen (siehe Abbildung 17-1). Dieses Vorgehen hat allerdings zwei Schwächen: Erstens, wie in Abbildung 17-1 des Beispiels *casual-leisure* zeigt, ist die größte Wahrscheinlichkeit jeweils die, dass auf den 50 ms Zeitabschnitt wiederum dieselbe Aktion folgt. Das ist darauf zurück zu führen, dass die Aktionen selbst immer mehrere Frames<sup>72</sup> lang sind und damit die Wahrscheinlichkeit, dass zwei gleiche Frames aufeinander folgen, am größten ist (wie die Grafik auch bestätigt).

Die zweite Schwäche ist die, dass nach [Kuhlthau, 1991] und den Ergebnissen der phasenweisen Klassifikation [siehe 11.3.7] diese Wahrscheinlichkeiten nicht über die ganze Suche konstant sind.

---

<sup>72</sup> Es wurde wiederum mit den 50ms Zeitfenstern gearbeitet.

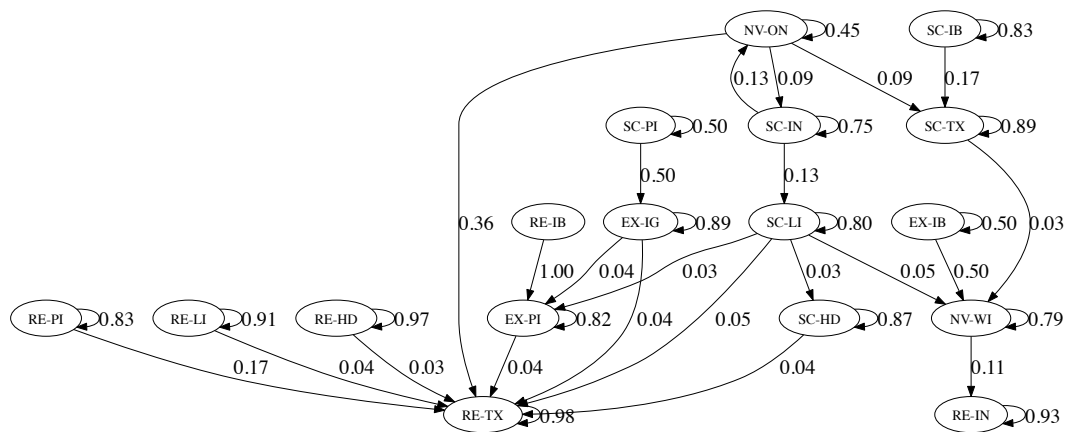


Abbildung 17-1 Übergangswahrscheinlichkeiten zwischen den Videolabels beim TT casual-leisure, Quelle [Knäusl und Ludwig, 2013]

Hier ist es demnach notwendig, eine andere Metrik für die Nutzeraktionen zu finden, um reelle Übergangswahrscheinlichkeiten berechnen zu können. Zusätzlich muss der Faktor, mit dem die Wahrscheinlichkeiten sich je nach Suchphase ändern, bekannt sein. Dazu sollte eine Verteilung der Nutzeraktionen in Abhängigkeit von den konkreten Suchphasen ermittelt werden.

### Verwendung von Hidden-Markov-Modellen

Das Hidden-Markov-Modell scheint geeignet, um die Übergangswahrscheinlichkeiten der verschiedenen Zustände respektive Phasen einer Suche zu berechnen, da diese nicht beobachtbar sind. Der „Vorteil“ eines HMM ist, dass die Zustände anhand bestimmter Emissionen angenommen werden, die in diesem Fall in Form der Nutzeraktionen vorliegen.

Qui [Qui, 1993] z.B. benutzt in einer Studie mit verschiedenen Nutzergruppen Markov-Modelle, um das Nutzerverhalten zu untersuchen. Die Nutzergruppen unterschieden sich dabei in Geschlecht, *search experience* und *search task*. Ziel dieser Studie war es, die Verteilung der sog. *search states* während der Suche in einem Hypertextsystem zu untersuchen. Die Tasks wurden unterschieden nach *specific task* vs. *general task*.

Mathematisch beschreibt [Qui, 1993] die Abfolge der Zustände als Markov Modell. Ziel ist hierbei nicht nur, die Wahrscheinlichkeiten für die einzelnen Zustände zu berechnen, sondern vor allem auch, mit welcher Wahrscheinlichkeit bestimmte Zustände

aufeinander folgen. [Qui, 1993] unterscheidet acht Zustände und erreicht mit einem HMM zweiter Ordnung gute Ergebnisse.

Yue, Han und He [Yue et al., 2013, Han et al., 2013] verwenden ebenfalls das Hidden-Markov-Modell (HMM) um das Nutzerverhalten im Information Retrieval zu untersuchen. Ihr Ziel ist es, sog. *search tactics* zu ermitteln. Diese werden explizit als die verborgenen Variablen modelliert, mit Hilfe des HMM untersuchen sie den Zusammenhang zwischen den einzelnen Aktionen des Nutzers und dessen *search tactics*. Ihre Ergebnisse zeigen, dass die so identifizierten *search tactics* wiederum konsistent sind mit dem *Information Seeking Process Modell* von Marchionini [Marchionini, 1995]. Die Autoren beklagen aber, dass in der bisherigen Forschung der Zusammenhang zwischen *user action* und *search tactics* zu wenig berücksichtigt wurde. Als *user action* definieren die Autoren:

Actions	Descriptions
Query [Q]	A user issues a query or clicks a query from search history
View [V]	A user click on a result in the returned result list
Save [S]	A user saves a snippet or bookmarks a webpage
Workspace [W]	A user clicks, edits or moments on an item saved in the workspace
Topic [T]	A user clicks on the topic statement for view or leaves comments

Die *search tactics* vergleichen die Autoren mit den *sub-processes* aus Marchioninis Modell, wobei sie einige *sub-processes* (z.B. *Define Problem*, *Select Source*, *Formulate Query*, *Examine Results* etc.) unter einer *search tactic* zusammenfassen. [Han et al., 2013] identifizierten so eine wahrscheinlichste Abfolge an *search tactics*. Eine Identifizierung der Zustände des Nutzers anhand der Aktionen bzw. seiner jeweiligen Vorgehensweise liefert Hinweise darauf, in welchem Stadium der Suche er sich befindet. Da die hier erhobenen Daten Hinweise enthielten, dass die verwendeten Inhalte u.A. auch abhängig vom Abschnitt der Suche sind, sind auch Kenntnisse über die Suchphase für eine adaptive Inhaltspräsentation notwendig. Die Untersuchungen von [Han et al., 2013; Yue et al., 2013 und Qui, 1993] liefern eine gute Basis für die Annahme, dass die *search tactics* und damit die Zustände des Nutzers auch entsprechend der Phasen, die [Kuhlt-

hau, 1991] identifiziert hat, und die Übergangswahrscheinlichkeiten mit Hidden-Markov-Modellen beschreibbar sind.

Problematisch allerdings könnte dabei sein, dass ein HMM von konstanten Wahrscheinlichkeiten ausgeht, diese sich aber möglicherweise – hier nicht nachweisbar – auch in Abhängigkeit der vorangegangenen Suchphasen ändern (siehe dazu Kapitel 12.4.1, Korrelation der Nutzeraktionen). Folglich müssten die Nutzeraktionen als Teil des Zustandes aufgenommen werden, was zu einem sehr komplexen Modell führen würde. Weiter geht ein HMM davon aus, dass der neue Zustand nur vom vorangegangenen abhängt – auch dies ist beim Suchprozess in der Wikipedia noch nicht ausreichend untersucht.

## 17.2 Weitere Aspekte zur adaptiven Inhaltspräsentation

Eine weitere Fragestellung, die sich bei der Frage einer Inhaltsgenerierung zur Laufzeit ergibt, ist, dass die inhaltlichen logischen Strukturen bei der Darstellung der Inhalte nicht verletzt werden dürfen. Die Aussagen der empirischen Untersuchungen dieser Arbeit beziehen sich auf die *Inhaltstypen*. Es ist aber weiter auch der Inhalt selbst entscheidend, da es sinnlos wäre, zwar das zu einem bestimmten Zeitpunkt das am stärksten präferierte Inhaltselement (z.B. Bild) zu präsentieren, wenn der inhaltliche und logische Zusammenhang zum Text und/oder den anderen soweit präsentierten Inhalten schwach oder nicht vorhanden ist bzw. zwingend notwendiges Vorwissen zuerst präsentiert werden muss. Das Selbe gilt für alle weiteren Inhaltselemente.

Knäusl und Ludwig [Knäusl und Ludwig, 2013] schlagen vor, die Inhalte unter Berücksichtigung der Konversationsmaximen von Grice [Grice, 1975] und den inhaltlichen Kohärenzbeziehungen [Grosz und Sidner, 1986] zu betrachten. Die wichtigsten Elemente sind hierbei:

- *satisfaction precedence* ( $x$  SP  $y$ ): Inhaltselement  $x$  *satisfaction precedes* Inhaltselement  $y$ :  $x$  sollte vor  $y$  präsentiert werden, da  $x$  Informationen enthält die notwendig sind, um  $y$  zu verstehen.



- *dominance* ( $x \text{ DOM } y$ ):  $x \text{ dominates } y$ , wenn  $y$  Inhalte enthält, die  $x$  ergänzen bzw. hilfreich sind, um  $x$  zu verstehen.

Die Menge der Inhaltstypen in den Wikipedia-Artikeln ist begrenzt. Ihre Zusammensetzung als Seite muss bestimmten Regeln folgen (nicht nur bei der adaptiven Präsentation). Somit kann man von einem begrenzten Inventar (Inhaltselemente) auf semantischer Ebene und von einem Bedarf nach Regeln zur Erstellung der Seite auf syntaktischer Ebene (Berücksichtigung von Kohärenzbeziehungen) sprechen. Diese Berücksichtigung bei der dynamischen Inhaltsgenerierung kann auch als Verfolgung der Konversationsmaximen von [Grice, 1975] betrachtet werden. Da die Maximen in erster Linie die Textgenerierung betreffen, hilft hier also die Analogie, dass die (verfügbaren) Inhaltselemente der Wikipedia und deren Zusammensetzung zu einer Seite auf der Basis einer speziellen Syntax durch die Auswahl des Nutzers erfolgt [Knäusl und Ludwig, 2013].

Dementsprechend müssen die Inhalte mit der Information, um welchen Typ es sich handelt, gekennzeichnet sein und darüber hinaus müssen die *satisfaction precedence* und die *dominance* Beziehungen der Elemente untereinander bekannt sein, so dass diese beim Auswählen der Inhalte berücksichtigt werden können.

Diese Eigenschaften der Inhalte beeinflussen zusätzlich zu den aus dem Nutzerkontext hergeleiteten Präferenzen die Präsentation der Inhalte. Die in dieser Arbeit gezeigten Zusammenhänge des Nutzungsszenarios mit dem Suchverhalten müssen der inhaltlichen und logischen Struktur der Inhalte untergeordnet werden.

### 17.3 Fehlende Daten im Kontextmodell

Im Modell sind Variablen enthalten, deren Daten nur mittelbar erfassbar sind (z.B. *Task Completion*, subjektive Relevanzbewertungen, Emotionswechsel während der Suche, Veränderung der Wissensstruktur).

Einige Daten können vermutet, geschätzt, oder interpoliert werden. [Moshfeghi und Jose, 2013] z.B. befragten ihre Probanden während des Experiments mehrfach zu emo-

tionalen Zuständen (Stress, Zuversicht etc.) und auch zu Einschätzungen der Inhalte (Überblick, Klarheit des Ziels etc.). Durch die erhaltenen Datenpunkte kann u.U. ein *Verlauf* dieser Zustände rekonstruiert werden. Es bleibt aber eine Fehlergefahr, da sich diese Zustände, bei entsprechendem Auslöser, sehr schnell ändern können. Darüber hinaus ist in einer Echtzeitanwendung nicht vorstellbar, den Nutzer mit häufigen Nachfragen zu unterbrechen. Auch im Labortest kann es die Validität der Daten beeinflussen, wenn der Nutzer mehrfach im Ablauf gestört wird (zum Beispiel für die Nachfrage, warum er gerade eine bestimmte Entscheidung getroffen hat).

Entsprechende Untersuchungen [Feild et al., 2010] zeigen, dass durch gewisse Hinweise in der Interaktion des Nutzers mit dem System auf Zustände wie Frustration geschlossen werden kann, letztendlich bleiben hier aber noch einige Fragen offen, die vor allem auch von der Persönlichkeit des Nutzers abhängen. Die Auswertungen zeigen, dass trotz Gemeinsamkeiten bei der Suche hinsichtlich der Definiertheit und der Endlichkeit des Suchziels die Nutzer auch eigene, persönliche Strategien haben.

Gleiches gilt für auch für die *Task Completion* [V<sub>11</sub>]. Selbst bei vorgegeben Suchzielen ist die Überprüfung ungleich schwierig, da erstens nicht im Detail (oder gar nicht) bekannt ist, welches Vorwissen der Nutzer jeweils hat und zweitens gerade bei weniger endlichen Suchzielen kein Messwert vorhanden ist, gegen den evaluiert werden kann. Möglicherweise wäre hier ein effektiverer Ansatz, die Stadien des ISP-Modells von [Kuhlthau, 1991] genauer zu operationalisieren, indem man untersucht, ob gewisse Verhaltensmuster (z.B. aus der Browserinteraktion aber auch mit Verfahren des Eye-trackings) den einzelnen Abschnitten zuzuordnen sind. Dann könnte zumindest eine Abschätzung erfolgen, an welchem Punkt der Suche sich der Nutzer hinsichtlich der *Task Completion* gerade befindet.

Ein weiterer wichtiger Aspekt, der zukünftig berücksichtigt werden sollte, ist die Tatsache, dass sich auch die initiativen Werte von  $S_1$ - $S_3$  und  $M$  während des Verlaufs unbewusst oder bewusst ändern können. Im Experiment wurde der Benutzer darauf hingewiesen, die Suche zu beenden, wenn er glaube, das Suchziel sei erfüllt, um die Wahrscheinlichkeit für diese Änderungen zu minimieren. Dennoch ist die Vermutung

vorhanden, dass diese Strategiewechsel, vor allem außerhalb der kontrollierten Laborsituation, *immer* vorhanden bleiben.

Das bedeutet in erster Linie, dass damit zu rechnen ist, dass die hier vorgegeben Suchziele sich abwechseln können, auch in hoher Frequenz. Das kann aus verschiedenen Gründen passieren, entweder weil äußerliche Bedingungen sich geändert haben (z.B. mehr oder weniger Zeit, Störfaktoren etc.) oder die inhaltlichen Ziele (neues Thema entdeckt, Serendipity-Effekt etc.). Es ist offen, wie schnell – sofern die persönlichen Präferenzen des Nutzers generell bekannt sind – ein System erkennen könnte, um welche Werte es sich bei  $S_1$  und  $S_2$  in etwa handelt.

## 17.4 Evaluation der Erkenntnisse dieser Arbeit

Die Erkenntnisse dieser Arbeit basieren auf den Theorien zu dieser Thematik in der Forschungsliteratur und der Interpretation der selbst gewonnenen Daten. Zur Planung einer konkreten situationsrelativen Informationspräsentation bzw. zur Inhaltsgenerierung zur Laufzeit in einem speziellen Anwendungskontext sollte sich zukünftige Forschung auch darauf konzentrieren, die in dieser Arbeit gewonnenen Erkenntnisse zu evaluieren.

Das Verhalten der Nutzer zeigte hier zwar unterscheidbare Präferenzen, allerdings wurde nicht untersucht, ob eine veränderte Präsentation, die bei den Suchszenarien in der Wikipedia die situativen Vorlieben berücksichtigt, von den Nutzer tatsächlich als besser bewertet wird.

## 17.5 Methodische Verbesserungen

### 17.5.1 Affektive Aktionen und Reaktionen

Die teilweise auftretenden methodischen Probleme dieser Arbeit können ebenfalls Gegenstand zukünftiger Forschung sein. Gerade im Bereich der Emotionen als Einfluss- und Zielgröße mangelt es zwar nicht an theoretisch vorhandener Methodik, durch die starke Verankerung der *casual-leisure* Suche im Alltag und den sehr persönlichen, teilweise affektiven Assoziationen mit den Inhalten aber ist die konkrete Erfassung affektiver Aktionen und Reaktionen schwierig.

Darüber hinaus ist die Suche nach Alltagsthemen in der Wikipedia wie sich hier zeigte, nicht so stark von Emotionen geprägt. In sensiblen Bereichen wie Krankheit, Sexualität oder Ähnlichem sind Einfluss und Bedeutung dieses Faktors vermutlich höher. Gerade dort wäre eine verbesserte Informationsrezeption erstrebenswert und zukünftige Forschung auf dem Gebiet der automatischen Informationserkennung und dem jeweiligen Einfluss dieser auf die Wünsche der Benutzer könnte sich hierauf konzentrieren.

### **17.5.2 Datenerhebung in Feldexperimenten**

Aus genannten Gründen wurden im Rahmen dieser Arbeit Laborexperimente durchgeführt. Darüber hinausgehende Informationen, wie die Informationssuche unterwegs, auf kleinen Geräten, im Freien, in Gruppen oder durch Störfaktoren beeinflusst etc., sind durch eine Feldstudie zu untersuchen.

### **17.5.3 Berücksichtigung von Kausalitäten**

In dieser Arbeit zeigte sich, dass gerade Bilder unterschiedlich verwendet bzw. berücksichtigt wurden. Es war bei der Berücksichtigung von Bildern Abhängigkeit von deren Funktion im Text und dem inhaltlichen Suchziel auszumachen. Bei der Studie zur Domain-abhängigen Relevanz von Inhaltselementen fiel weiter auf, dass die VPs die Bilder, die Porträts der im Artikel behandelten Person enthielten, nur in den wenigsten Fällen berücksichtigt haben. Hier sind gezielte Untersuchungen zu den Kausalitäten für diese Entscheidungen vorstellbar: Warum werden diese Bilder im Einzelfall berücksichtigt und warum nicht? Sind dafür personenbezogene Faktoren (Bildinhalt bereits bekannt etc.) verantwortlich oder die Platzierung auf der Seite? Diese Einflüsse sind auch bei weiteren Informationselementen denkbar.

### **17.5.4 Weitere Untersuchungsgegenstände**

Die empirischen Untersuchungen dieser Arbeit fanden ausschließlich an der deutschsprachigen Wikipedia statt. Die grundsätzliche Fragestellung bezog sich ebenfalls vorrangig auf dieses Informationssystem.

Für zukünftige Forschungen bezüglich der nutzer- und situationsrelativen Informationspräsentation könnten weitere Informationssysteme aus verschiedenen Bereichen fokussiert und der Nutzen einer adaptiven Inhaltspräsentation unter Berücksichtigung des hier definierten und zusammengestellten Nutzerkontextes stattfinden.

## 18 Verzeichnisse

### 18.1 Abbildungsverzeichnis

Abbildung 5-1 Standardmodell des Suchprozesses von [Hearst, 2009, Kapitel drei].....	45
Abbildung 5-2 Information Behaviour Model von [Wilson, 1999, Kapitel 2] .....	48
Abbildung 5-3 The cognitive communication system for Information Science, information seeking and IR. Revision of Ingwersen [1992, p.33; 1996, p.6], from [Belkin, 1978], [Ingwersen und Järvelin, 2005, S.39] .....	49
Abbildung 5-4 Interactive Information Seeking, Retrieval and Behavioral processes [Ingwersen und Järvelin, 2005, S.277] .....	50
Abbildung 5-5 Information Seeking Model von Byström und Järvelin [Byström und Järvelin, 1995, S.7] .....	51
Abbildung 5-6 Head und Eisenberg, 2010, Ziele der Wikipedianutzung bei einer Umfrage bei College-Studenten 2008.....	59
Abbildung 5-7 Auszug aus der Studie von [Lim, 2009] zur Wikipedia Nutzung - cognitive needs .....	60
Abbildung 5-8 Auszug aus der Diary Study von [Elsweiler et al., 2011, S. 10] .....	63
Abbildung 5-9 Auszug aus der Twitter-Studie von [Elsweiler et al., 2011, S.12] .....	64
Abbildung 6-1 Task-Komplexitäts Klassifikation nach [Byström und Järvelin, 1995], Quelle: [Ingwersen und Järvelin, 2005, S.75] .....	81
Abbildung 6-2 Schematische Darstellung des Kontextmodells.....	85
Abbildung 7-1 Überblick über ein adaptive gestaltetes Information Retrieval System [Lau et al., 2008, S.4] .....	100
Abbildung 9-1 Infobox des Wikipedia-Artikel über Regensburg, abgerufen Mai 2013 von <a href="http://de.wikipedia.org/wiki/Regensburg">http://de.wikipedia.org/wiki/Regensburg</a> .....	115
Abbildung 9-2 Monatliche Top 100 Wikipedia-Artikel Sept. 2006 - Januar 2007 aus [Spoerri, 2007] .....	119
Abbildung 9-3 Kategorisierung der Artikelstichproben.....	130
Abbildung 9-4 Verteilung des Bereichs Gegenstands "Typ" eines Artikels der Wikipedia-Stichprobe.....	130
Abbildung 9-5 Verteilung des Bereichs "Typ" der Stichprobe aus der Brockhaus Enzyklopädie .....	132

Abbildung 9-6 Verteilung der Artikel der Stichproben auf die Themenbereiche, in %, blau= WP, rot=BE.....	134
Abbildung 10-1 Anatomie einer Katze, Uwe Gille, Quelle: <a href="http://upload.wikimedia.org/wikipedia/commons/8/8b/Scheme_cat_anatomy-de.svg">http://upload.wikimedia.org/wikipedia/commons/8/8b/Scheme_cat_anatomy-de.svg</a> .....	137
Abbildung 10-2 schlafende Babykatze, Quelle: <a href="http://upload.wikimedia.org/wikipedia/commons/f/fb/Sleeping_baby_cat.jpg">http://upload.wikimedia.org/wikipedia/commons/f/fb/Sleeping_baby_cat.jpg</a> .....	137
Abbildung 10-3 Bild aus der Studie: Mozart and Linley 1770, Quelle: <a href="http://de.wikipedia.org/wiki/Mozart#mediaviewer/Datei:Mozart_and_Linley_1770.jpg">http://de.wikipedia.org/wiki/Mozart#mediaviewer/Datei:Mozart_and_Linley_1770.jpg</a> .....	151
Abbildung 11-1 Ausschnitt aus Wikipedia: Regensburg, abgerufen 13. Dezember 2013.....	160
Abbildung 11-2 Global Structure of emotion types, [Ortony1988, S.19].....	169
Abbildung 11-3 Circumplexmodell [Russell, 1980, S.7].....	170
Abbildung 11-4 Circumplexmodell [Desmet und Hekkert, 2007], basierend auf [Russell, 1980].....	171
Abbildung 11-5 The Self Assessment Manikin (SAM), [Bradley und Lang, 1994, S. 3].....	175
Abbildung 11-6 Klassifikation der Gesichtsbewegungen nach Ekman's Facial Action Coding System [Ekman und Rosenberg, 1997], Quelle: <a href="http://www.gnosisfacialis.de/Seiten/facs/facs1.htm">[http://www.gnosisfacialis.de/Seiten/facs/facs1.htm]</a> abgerufen am 17. September 2013.....	177
Abbildung 11-7 Overview of facial muscles [source: <a href="http://www.musashi-rie.com/e-kaokin.htm">http://www.musashi-rie.com/e-kaokin.htm</a> ], abgerufen am 15. Juni 2012.....	178
Abbildung 11-8 [Picard, 1998, S.4] Signale des EMG im Vergleich mit Puls und Leitfähigkeit der Haut.....	179
Abbildung 11-9 Datenbeispiel Browserlogs.....	184
Abbildung 11-10 Laboraufbau Studie I und II.....	191
Abbildung 11-11 Angaben der Nutzer, wie realistisch sie die Aufgabenstellung einschätzte....	196
Abbildung 11-12 Nutzung der Wikipedia bei den Probanden.....	198
Abbildung 11-13 Software "Anvil" zur Annotation der Videos.....	198
Abbildung 11-14 Mittelwerte der Häufigkeiten der Videolabel-Kombinationen (B= casual leisure, L = Learn, LU = look up).....	201
Abbildung 11-15 Aufsummierung aller nach Kruskal-Wallis-Test statistisch signifikant unterschiedlich verteilten Kombinationen (Mittelwerte, absolute Häufigkeiten).....	202
Abbildung 11-17 Mittelwerte der relativen Häufigkeiten der Videolabel-Kombinationen, alle TT, alle VPs.....	209
Abbildung 11-18 Balkendiagramm der Wahrscheinlichkeiten/Nutzeraktionen.....	213
Abbildung 11-19 Histogramm BE-BI Häufigkeit der relativen Häufigkeiten, alle drei TT .....	214
Abbildung 11-20 Histogramm LE-AA Häufigkeit der relativen Häufigkeiten, alle drei TT.....	215
Abbildung 11-21 Histogramm QL-LI Häufigkeit der relativen Häufigkeiten, alle drei TT .....	216

Abbildung 11-22 Matrix der drei am unterschiedlichsten Verteilten Nutzeraktionen, relative Häufigkeiten.....	218
Abbildung 11-23 Summe der absoluten Häufigkeiten der Videolabel-Kombinationen (50ms frames) aller VPs .....	224
Abbildung 11-24 emotionale Selbsteinschätzung der Probanden anhand des im Fragebogen verwendeten SAM. (arousal: 1 = erregt, 5 = ruhig, valenz: 1 = angenehm, 5 = unangenehm) ..	227
Abbildung 12-1 Kruskal-Wallis-Test über die Verteilung von NA und IN über alle drei Tasktypen .....	244
Abbildung 12-2 Histogramm der Angaben zur Selbsteinschätzung der emotionalen Lage mittels SAM über alle tasks (valence 1=angenehm, 5= unangenehm, arousal 1= erregt, 5 = ruhig) .....	247
Abbildung 12-3 relative Häufigkeiten EX-PI Verteilung alle drei Tasktypen Boxplot .....	262
Abbildung 12-4 relative Häufigkeiten RE-TX Verteilung alle drei Tasktypen Boxplot,.....	262
Abbildung 12-5 relative Häufigkeiten RE-IX Verteilung alle drei Tasktypen Boxplot.....	263
Abbildung 12-6 relative Häufigkeiten RE-IN Verteilung alle drei Tasktypen Boxplot.....	263
Abbildung 12-7 Ausschnitt aus dem Artikel über die Bayerische Zugspitzbahn in der deutschsprachigen Wikipedia [abgerufen: 12.Juli 2013, Quelle: <a href="http://de.wikipedia.org/wiki/Bayerische_Zugspitzbahn">http://de.wikipedia.org/wiki/Bayerische_Zugspitzbahn</a> ] .....	269
Abbildung 12-8 Zuordnung der Untersuchungsgegenstände zur Klassifikation, [Vogel, 2013, s.34] .....	271
Abbildung 12-9 Dauer der Betrachtung der Inhaltselemente Thema A und Thema B Summe aller VPs.....	273
Abbildung 12-10 Dauer der Betrachtung der Inhaltselemente Thema A und Thema B Summe aller VPs ohne TX .....	274
Abbildung 12-11 Fixationsdauer des Inhaltelements Bild in Summe aller VPs [Vogel, 2013, S. 51] .....	274
Abbildung 17-1 Übergangswahrscheinlichkeiten zwischen den Videolabels beim TT casual-leisure, Quelle [Knäusl und Ludwig, 2013].....	298
Abbildung 20-1 relative Häufigkeitgen Taskweise ( $n_{\text{Nutzeraktion}}/n_{\text{Summe}}$ ).....	389

## 18.2 Tabellenverzeichnis

Tabelle 6-1 Modell des Nutzers und des Informationsbedürfnisses.....	77
Tabelle 6-2Modell des weiteren Kontext .....	82
Tabelle 6-3 Modell des Suchverhaltens .....	83
Tabelle 9-1 Kategorien mit den meisten Zuordnungen in der Deutschsprachigen Wikipedia .	123

Tabelle 9-2 Kategorie "Typ" .....	126
Tabelle 9-3 Kategorie "Bereich" .....	127
Tabelle 9-4 Kategorie "Zeitraum" .....	128
Tabelle 9-5 Kategorie „Regionalität“ .....	129
Tabelle 9-6 Verteilung der Artikel auf die verschiedenen Typen, WP und BE in % .....	131
Tabelle 9-7 Verteilung der Kombination „Person“ + Themenbereiche WP und BE in % .....	133
Tabelle 9-8 Zuordnung der Artikel zu den Bereichen Wikipedia und BE .....	135
Tabelle 10-1 Bildauswahl Häufigkeiten Mozarttest .....	152
Tabelle 10-2 Durchschnittliche Bewertung der Bilder mit Schulnoten .....	153
Tabelle 10-3 Auswahl der Begründungen Fall 1. ....	154
Tabelle 10-4 Auswahl der Begründungen Fall 2. ....	154
Tabelle 10-5 Auswahl der Begründungen bei allen Bildern .....	155
Tabelle 10-6 Ergebnisse aller drei Ereignisse .....	156
Tabelle 11-1 Zusammenstellung von Emotionen nach [Kleinginna und Kleinginna, 1981, S.17] .....	163
Tabelle 11-2 Videolabels "Aktion" .....	181
Tabelle 11-3 Videolabels "Inhaltselement" .....	182
Tabelle 11-4 Mögliche Kombinationen der Videolabels .....	183
Tabelle 11-5 im Experiment vorgegebene Variablen .....	188
Tabelle 11-6 Im Experiment zu erfassende Daten aus dem Modell .....	189
Tabelle 11-7 weitere zu erfassende Variablen aus dem Modell .....	189
Tabelle 11-8 Datenbeispiel Videolabels/Nutzeraktionen .....	200
Tabelle 11-9 Datenbeispiel Häufigkeiten Nutzeraktionen/Task .....	200
Tabelle 11-10 Ergebnisse des Kruskal Wallis Test über alle drei TT, absolute Häufigkeiten .....	204
Tabelle 11-11 Games - Howell Mehrfachvergleich, absolute Häufigkeiten .....	205
Tabelle 11-12 Chi-Quadrat Test über die Verteilung der Nutzeraktionen in den drei Tasktypen bei sechs VPs, absolute Häufigkeiten [Knäusl et al., 2012] .....	208
Tabelle 11-13 Kruskal-Wallis-Test, Verteilung der $h_n$ über alle TT .....	210
Tabelle 11-14 Games-Howell-Test, Mehrfachvergleich, $h_n$ .....	211
Tabelle 11-15 Zusammenfassung der Test-Ergebnisse. Spalte 1-3: Ergebnisse des t-Tests, unter Verwendung der absoluten (n) und der relativen Häufigkeiten ( $h_n$ ). Spalte 5&6: Ergebnisse des Kruskal-Wallis-Tests, unter Verwendung der relativen Häufigkeiten ( $h_n$ ) .....	217
Tabelle 11-16 Ergebnisse der automatischen Klassifikation unter Verwendung nur einer Videolabel-Kombination .....	220
Tabelle 11-17 [Bradley und Lang, 1994, S.50] Faktorenladung der Wortpaare des Semantischen Differential nach Reduktion auf die drei Bereiche "Valence", "Arousal" und "Dominance" .....	225



Tabelle 11-18 Korrelation nach Pearson Erfolg und Zufriedenheit mit der Suche .....	228
Tabelle 11-19 Deskriptive Statistik zur Angabe "thematische Vorkenntnis" .....	230
Tabelle 11-20 binär logistische Regression learn .....	231
Tabelle 11-21 binär logistische Regression look up.....	232
Tabelle 11-22 binär logistische Regression casual - leisure.....	233
Tabelle 11-23 Test auf Normalverteilung der Angaben in den Dimensionen arousal und valence.....	237
Tabelle 12-1 Videolabels "Inhaltselement" .....	241
Tabelle 12-2 Mögliche Kombinationen der Videolabels .....	241
Tabelle 12-3 Unterscheidung in navigierende und informierende Labelkombinationen .....	243
Tabelle 12-4 Verhältnis von IN zu NA bei allen drei Tasktypen.....	244
Tabelle 12-5 Test auf Normalverteilung der Angaben mittels SAM.....	247
Tabelle 12-6 Lageparameter und Häufigkeiten valence und arousal.....	248
Tabelle 12-7 t-Test "Mehr Foto/Text/Grafiken gewünscht" der Gruppen positiv "arousal" und negativ „arousal“ (Trennwert 4).....	250
Tabelle 12-8 t-Test "Mehr Foto/Text/Grafiken gewünscht" der Gruppen positiv "valence" und negativ "valence" (Trennwert 4) .....	250
Tabelle 12-9 Mehrfachvergleich der Aufgabentypen auf unterschiedliche Wünsche.....	252
Tabelle 12-10 Mehrfachvergleich für "Element in dem Info gefunden wurde" .....	254
Tabelle 12-11 t-Test Zufriedenheit mit dem Suchverlauf und dem Wunsch nach einem bestimmten Inhaltstyp.....	256
Tabelle 12-12 Korrelation zwischen der Zufriedenheit mit dem Suchverlauf und angemessener Präsentation .....	256
Tabelle 12-13 Korrelation zwischen dem Sucherfolg und der Bewertung, ob die Präsentation der Inhalte angemessen war .....	257
Tabelle 12-14 Kruskal-Wallis-Test über die Verteilung aller Videolabel-Kombinationen über alle drei TT .....	259
Tabelle 12-15 Mehrfachvergleich (Games-Howell) der unterschiedlich verteilten Nutzeraktionen.....	260
Tabelle 12-16 Verteilung der Betrachtungsdauer auf die AoIs Thema A und Thema B und Chi-Quadrat-Verteilungstest über Summe und Dauer .....	272
Tabelle 13-1 Einfluss der Variablen des Modells.....	277
Tabelle 20-1 Aufgaben Studie eins SS 2012 .....	347
Tabelle 20-2 Videolabels extrahiert P02-screcc-annotiert-LU (Look-up Task) Studie 1.....	354
Tabelle 20-3 Absolute Häufigkeiten alle Tasks Studie eins .....	355
Tabelle 20-4 Test auf Normalverteilung der Häufigkeiten über alle Tasks.....	359

Tabelle 20-5 Test auf Normaverteilung Häufigkeiten casual leisure .....	359
Tabelle 20-6 Test auf Normalverteilung Häufigkeiten learn.....	359
Tabelle 20-7 Test auf Normalverteilung Häufigkeiten look up .....	360
Tabelle 20-8 Test auf Varianzgleichheit zwischen den Gruppen .....	360
Tabelle 20-9 absolute Häufigkeiten casual-leisure.....	361
Tabelle 20-10 absolute Häufigkeiten learn .....	361
Tabelle 20-11 absolute Häufigkeiten look up .....	361
Tabelle 20-12 relative Häufigkeiten casual-leisure .....	362
Tabelle 20-13 relative Häufigkeiten learn .....	362
Tabelle 20-14 relative Häufigkeiten look up.....	363
Tabelle 20-15 Mittelwerte der relativen Häufigkeiten für die Videolabelkombinationen, alle TT .....	363
Tabelle 20-16 T-test über Kategorie arousal des SAM, Trennwert 4, $n < 4 = 41$ , $n > 4 = 48$ .....	363
Tabelle 20-17 T-test über Kategorie valences des SAM, Trennwert 3, $n \leq 3 = 64$ , $n > 3 = 25$ ....	365
Tabelle 20-18 Test über die Gleichverteilung der Videolabelkombinationen in den ersten 50 x 50 ms, absolute Häufigkeiten .....	366
Tabelle 20-19 t-Test Männlich vs. weiblich VP, relative Häufigkeiten.....	367
Tabelle 20-20 t-Test Gruppe mit keinem oder wenig Vorwissen vs. Vorwissen .....	368
Tabelle 20-21 Korrelation der Videolabels untereinander, relative Häufigkeiten .....	369
Tabelle 20-22 Aufgaben Studie WS 2012/2013 .....	371
Tabelle 20-23 deskriptive Statistik post Task Fragebogen.....	375
Tabelle 20-24 Test auf Normalverteilung Angaben Post Task Fragebogen.....	376
Tabelle 20-25 Videolabels/Nutzeraktionen deskriptive Statistik.....	376
Tabelle 20-26 Test auf Normalverteilung relative Häufigkeiten look up .....	386
Tabelle 20-27 Test auf Normalverteilung relative Häufigkeiten casual leisure .....	386
Tabelle 20-28 Test auf Normalverteilung, relative Häufigkeiten learn.....	387
Tabelle 20-29 Test auf Normalverteilung über alle Gruppen, relative Häufigkeiten .....	387
Tabelle 20-30 Test auf Homogenität der Varianzen alle drei TT, relative Häufigkeiten .....	388
Tabelle 20-31 Relative Werte Studie zwei ( $n_{\text{Nutzeraktion}}/n_{\text{Summe}}$ ).....	388
Tabelle 20-32 Vergleich der Look-up Tasks Thema 1 vs. Thema 2, t-Test, relative Häufigkeiten .....	390
Tabelle 20-33 t-Test Gruppe arousal positiv und arousal negativ, Trennwert 4, Wahrscheinlichkeiten der Videolabelkombinationen .....	391
Tabelle 20-34 t-Test Gruppe valence positiv und valence negativ, Trennwert 4, relative Häufigkeiten der Videolabelkombinationen.....	392

---

Tabelle 20-35 t-Test Gruppen viel und wenig WP-Kenntnis, Trennwert 2, relative Häufigkeiten der Videolabelkombinationen .....	393
Tabelle 20-36 t-Test zw. Gruppen viel und wenig thematische Vorkenntnis, Trennwert 2.....	394
Tabelle 20-37 Mehrfachvergleich zwischen den Aufgabentypen, Häufigkeiten .....	395
Tabelle 20-38 T-Test über die Mittelwerte aller drei TT emotionale Lage positiv vs. negativ (summe valence und arousal), absolute Häufigkeiten .....	396
Tabelle 20-39 Korrelation nach Pearson der relativen Wahrscheinlichkeiten der Videolabels.	397

### 18.3 Abkürzungsverzeichnis

IR	Information Retrieval
BE	Brockhaus Enzyklopädie
nibs	Nicht informationsbezogenes Suchziel
ASK	anamolous state of knowledge
VP	Versuchsperson
KI	Künstliche Intelligenz
SAM	self assessment manikin
facs	facial action coding system
TT	Tasktyp
AoI	Area of Interest
L	Tasktyp look up
LE	Tasktyp learn
B	Tasktyp causal-leisure
HMM	Hidden Markov Model
SPSS	Marke der Softwarefirma IBM (ursprünglich: <i>Statistical Package for the Social Sciences.</i> )

## 19 Literaturverzeichnis

- Adafre, S. F. & de Rijke, M. (2006), Exploratory Search in Wikipedia, in 'Proceedings of Special Interest Group on Information Retrieval 2006, Workshop on Evaluating Exploratory Search Systems'.
- Adfinger, M.; Schneider, M. & Lechler, D. (2013), 'Untersuchung der Augenbewegung beim Information Retrieval', Seminararbeit zum Thema situative Relevanz, Universität Regensburg, Lehrstuhl für Informationswissenschaft.
- Agichtein, E.; Brill, E. & Dumais, S. (2006), Improving Web Search Ranking by Incorporating User Behavior Information, in 'Proceedings of Special Interest Group on Information Retrieval 2006', ACM, New York, NY, USA, pp. 19 - 26.
- Agichtein, E.; Brill, E.; Dumais, S. & Ragno, R. (2006), Learning user interaction models for predicting web search result preferences, in 'Proceedings of SIGIR', pp. 3-10.
- Alexa (2013), 'Alexa, the web information company', online verfügbar unter [<http://www.alexa.com/>], abgerufen September 2013.
- Allen, B. (1991), 'Topic Knowledge and Online Catalog Search Formulation', Allen, B. (1991). Topic Knowledge and Online Catalog Search Formulation. The Library Quarterly **61**(2), pp. 188-213.
- Anthony, D.; Smith, S. W. & Williamson, T. (2007), 'The Quality of Open Source Production: Zealots and Good Samaritans in the Case of Wikipedia', Technical Report, Hanover.
- Arapakis, I.; Jose, J. M. & Gray, P. (2008), Affective Feedback: An Investigation into the Role of Emotions in the Information Seeking Process, in 'Proceedings of the Special Interest Group on Information Retrieval 2008', pp. 395-402.

- Arapakis, I.; Konstas, I. & Jose, J. M. (2009), Using facial Expressions and Peripheral Physiological Signals as Implicit Indicators of Topical Relevance, in 'Proceedings of the 17th ACM international conference on Multimedia', pp. 461 - 470.
- Arazy, O.; Morgen, W. & Patterson, R. (2006), 'Wisdom of the Crowds: Decentralized Knowledge Construction in Wikipedia', Social Science Research network 16th Annual Workshop on Information Technologies & Systems (WITS) Paper.
- Athitsos, V.; L Frankel, C. & Swain, M. J. (1997), Distinguishing photographs and graphics on the World Wide Web, in 'CAIVL '97 Proceedings of the 1997 Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '97)', p. 10.
- Athitsos, V.; L Frankel, C. & Swain, M. J. (1996), 'WebSeer: An Image Search Engine for the World Wide Web', Technical Report, University of Chicago Chicago, IL, USA.
- Axelrod, L. (2004), The affective connection: how and when users communicate emotion, in 'CHI '04 Extended Abstracts on Human Factors in Computing Systems', ACM, New York, NY, USA, pp. 1033-1034.
- Bär, J. (2008), 'Diplomarbeit: Analyse zur Automatischen Klassifizierung von Bildern', TU Chemnitz, Fakultät für Informatik, Professur für Medieninformatik.
- Ballstaedt, S.-P. (2005), 'Fachhochschule Gelsenkirchen: Visualisierung: Bilder in der technischen Kommunikation', Vorlesungsskript.
- Baltrunas, L.; Kaminskas, M.; Ludwig, B.; Moling, O.; Ricci, F.; Aydin, A.; Lüke, K.-H. & Schwaiger, R. Huemer, C. & Setzer, T., ed., (2011), InCarMusic: Context-Aware Music Recommendations in a Car in: E-Commerce and Web Technologies, Springer Verlag Berlin Heidelberg, pp. 89-100.
- Barthes, R. (1964), 'Rhétorique de l'image.', Communications 1964(4), pp. 40-51.

- Bates, M. (1989), 'The design of browsing and berrypicking techniques for the online search interface', *Online Review* 13(5), pp. 407-424.
- Beitzel, S.; Jensen, E.; Frieder, O.; Grossman, D.; Lewis, D.; Chowdhury, A. & Kolcz, A. (2005), Automatic Web Query Classification Using Labeled and Unlabeled Training Data, in 'Proceedings of Special Interest Group on Information Retrieval 2005', ACM, New York and NY, pp. 581-582.
- Belkin, N. (1993), 'Interaction with Texts: Information Retrieval as Informatio-Seeking Behavior', *Information Processing and Management* 31(3), pp. 431-448.
- Belkin, N. J. (2006), 'Getting personal: Personlization of support for interaction with information', Talk at The 1st International Workshop on Adaptive Information Retrieval.
- Belkin, N. J. (1980), 'Anomalous states of knowledge as a basis for information retrieval', *The Canadian Journal of Information Science* 5, pp.133-143.
- Belkin, N. J. (1977), 'A concept of information vor information science', Unpublished doctoral dissertation, University College, University of London.
- Berger, H.; Dittenbach, M. & Merkl, D. (2004), An adaptive information retrieval system based on associative networks, in 'Proceedings of the first Asian-Pacific conference on Conceptual modelling - Volume 31', Australian Computer Society, Inc., Darlinghurst, Australia, pp. 27-36.
- Bilal, D. (2002), 'Perspective on children's navigation of the World Wide Web: Does the type of search task make a different?', *Online Information Review* 26(2), pp.108 - 117.

- Bilal, D. & Kirby, J. (2002), 'Differences and similarities in information seeking: children and adults as Web users', *Information Processing and Management* 38(5), pp. 649-670.
- Blair, D. C. (1990), 'Language and representation in information retrieval', Elsevier Science Ltd.
- Bockes, F.; Bruckmaier, K.; Covaci, S.; Gottswinter, B. & Scherr, A. (2013), 'Informationssuche mit Wikipedia - Erfahrungssache?', Seminararbeit WS 2012/2013, Universität Regensburg, Lehrstuhl für Informationswissenschaft.
- Borlund, P. (2003), 'The IIR evaluation model: a framework for evaluation of interactive information retrieval systems', *Information Research* 8(3), pp. 1-16.
- Borlund, P. (2000), *Evaluation of interactive information retrieval systems*, Abo Akademi University Press.
- Borlund, P. & Ingwersen, P. (1997), 'The Development of a Method for the Evaluation of Interactive Information Retrieval Systems', *Journal of Documentation* 53(3), pp. 225-251.
- Borlund, P. & Schneider, J. (2010), Reconsideration of the Simulated Work Task Situation: A Context Instrument for Evaluation of Information Retrieval Interaction, in 'Proceedings of the IiiX 2010', ACM Press, pp. 154 - 164 .
- Bortz, J. & Döring, N. (2001), *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler*, Springer Verlag Berlin Heidelberg.
- Bradley, M. M. & Lang, P. J. (1994), Measuring emotion: The self-assessment manikin and the semantic differential, 'Journal of Behavior Therapy and Experimental Psychiatry' 25(1), pp. 49-59.



- Brand-Gruwel, S.; Wopereis, I. & Vermetten, Y. (2005), 'Information problem solving by experts and novices: Analysis of a complex cognitive skill', *Computers in Human Behavior* 21, pp. 487-508.
- Broder, A. (2002), 'A taxonomy of web search', *ACM SIGIR Forum* 36(2), pp. 3-10.
- Brookes, B. (1980), 'The foundations of information science, Part 1. Philosophical aspects', *Journal of the Information Science* 2(3/4), pp. 125-134.
- Brown, S. L. & Schwartz, G. E. (1980), 'Relationships between facial Electromyography and subjective experience during affective imagery', *Biological Psychology* 11, pp. 49 - 62.
- Brusilovsky, P.; Stock, O. & Stapparava, C. Brusilovsky, P.; Stock, O. & Stapparava, C., ed., (2000), *Adaptive Hypermedia and Adaptive Web-Based Systems*, Springer Verlag, Chapter: Preface.
- Buckland, M. (1991), 'Information as a Thing', *Journal of the American Society for Information Science* 42(5), pp. 351-360.
- Buscher, G.; Dengel, A.; Biedert, R. & Elst, L. V. (2012), 'Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond', *ACM Trans. Interact. Intell. Syst.* 1(2), pp- 9:1-9:30.
- Buscher, G.; Dengel, A. & Van Elst, L. (2008), Eye movements as implicit relevance feedback, in *Proceedings of the international conference of Human-Computer Interaction 2008: Extended Abstracts on Human Factors in Computing Systems*, pp. 2991-2996.
- Bush, V. (1945), 'As we may think', *The Atlantic Monthly*, July 1945, pp. 101- 108.

- Byström, K. & Järvelin, K. (1995), 'Task complexity affects information seeking and use', *Information Processing & Management* 32(2), pp. 191 - 213.
- Cacioppo, J. T. e. a. (1986), 'Electromyographic Activity Over Facial Muscle Regions Can Differentiate the Valence and Intensity of Affective Reactions', *Journal of Personality and Social Psychology* 50(2), pp. 260 - 268.
- Cai, Y.; Tong, W.; Yang, L. & Hauptmann, A. G. (2012), Constrained keypoint quantization: towards better bag-of-words model for large-scale multimedia retrieval, in 'Proceedings of the 2nd ACM International Conference on Multimedia Retrieval', ACM, New York, NY, USA, pp. 16:1-16:8.
- Calderon-Benavides, L.; Gonzalez-Caro, C. & Beza-Yates, R. (2010), Towards a Deeper Understanding of the User's Query Intent, in 'Proceedings of Special Interest Group on Information Retrieval 2010, Workshop on Query Representation and Understanding', ACM Press, Geneva, Switzerland, pp. 21-24.
- Cao, H.; Jiang, D.; Pei, J.; Chen, E. & Li, H. (2009), Towards Context-Aware Search by Learning A Very Large Variable Length Hidden Markov Model from Search Logs, in 'Proceedings of the International World Wide Web Conference 2009', ACM, New York and NY, pp. 191-201.
- Capurro, R. (2000), 'Einführung in den Informationsbegriff: Der Informationsbegriff in der Informationswissenschaft'. Online verfügbar unter: [<http://www.capurro.de/infovorl-kap2.htm>, abgerufen 23.April 2013]
- Stegbauer, C. (2009), Wikipedia. Das Rätsel der Kooperation, VS Verl. für Sozialwissenschaften.
- Clark, M.; Ruthven, I.; Holt, P. & Song, D. (2012), Looking for Genre: the use of Structural Features During Search Tasks with Wikipedia, in 'Proceedings of the IiiX 2012, Nijmegen, The Netherlands', pp. 145 - 154.

- Claypool, M.; Le, P.; Wased, M. & Brown, D. (2001), Implicit interest indicators, in 'Proceedings of the International Conference on Intelligence User Interfaces 2001', pp. 33-40.
- Cohen, J. (1988), Statistical power analysis for the behavioral sciences, Erlbaum, Hillsdale.
- Crestani, F. & Rijsbergen, C. J. V. (1997), 'A Model for Adaptive Information Retrieval', Journal Intelligence Information Systems 8(1), pp. 29-56.
- Cutrell, E. & Guan, Z. (2007), What are you looking for? An eye-tracking study of information usage in web search, in 'Proceedings of the SIGCHI conference on Human factors in computing systems', ACM, New York, NY, USA, pp. 407-416.
- D., B.; Kühme, T.; Malinowski, U. & Sukaviriya, P. (Januar 1994), Computer-aided Adaption of User Interfaces, in 'SIGCHI Bulletin,' 26(1), pp. 25 - 27.
- Dahm, M. (2006), Grundlagen der Mensch-Computer-Interaktion, Pearson Studium, München.
- Darwin, C.; Beer, U., ed. (1965), Der Ausdruck der Gefühle bei Mensch und Tier, Nach d. Übers. von Theodor Bergfeldt, Rau, Düsseldorf.
- Debowski, S. (2001), 'Wrong way: Go back! An exploration of novice search behaviours while conducting an information search', The Electronic Library 19, pp. 371-382.
- Dervin, B. in Glazier, J. & Powell, R.R., ed., (1992), From the mind's eye of the 'user': The sense-making qualitative-quantitative methodology, Englewood, CO: Libraries Unlimited.

- Desmet, P.; Schifferstein, H. & Hekkert, P., ed., (2007), *Product emotion*, Elsevier Science Ltd.
- Duchowski, A. T. (2007), *Eye Tracking Methodology*, Springer Verlag, London.
- Ekman, P. (1992), 'An Argument for Basic Emotion', In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion*, Sussex, UK: John Wiley & Sons, Ltd.
- Ekman, P. & Rosenberg, Erika, L., ed. (1997), *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Oxford University Press
- Elsweiler, D. & Wilson, M. (2010), *Casual-leisure Searching: the Exploratory Search scenarios that break our current models*, in 'Proceedings of the 2010 Symposium on Human-Computer Interaction and Information Retrieval'.
- Elsweiler, D.; Wilson, M. & Harvey, M. (2010), 'Searching for Fun: Casual-leisure Search', *Proceedings of the European Conference of Information Retrieval 2012*, Barcelona, Spain
- Elsweiler, D.; Wilson, M. L. & Kirkegaard Lunn, B. (2011), *New Directions in Information Behaviour*, Emerald Publishing, chapter: *Understanding Casual-leisure Information Behaviour*, pp. 211-239.
- Feild, H. A.; Allan, J. & Jones, R. (2010), *Predicting Searcher Frustration*, in 'Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 34-41.
- Felder, R. (1993), 'Reaching the Second Zier: Learning and Teaching Styles in College Science Education', *J. College Science Teaching* 23(5), pp. 286-290.

- Fischer, K.; Erdelez, S. & McKechnie, L. (2005), *Theories of Information Behavior*, Information Today, Inc., 2005, Medford, New Jersey.
- Fonseca, B.; Golgher, P.; Possas, B.; Ribeiro-Neto, B. & Ziviani, N. (2005), *Cocept-Based Interactive Query Expansion*, in 'Proceedings of international Conference on Information an Knowledge Management 2005', ACM, New York, NY, USA, pp. 696 - 703.
- Forster, S.; Lamm, L. & Semmelmann, T. (2013), 'Studie zum Leseverhalten für nutzer- und situationsadaptive Informationssysteme', Seminararbeit WS 2012/2013, Universtität Regensburg, Lehrstuhl für Informationswissenschaft.
- Fox, S.; Karnawat, K.; Mydland, M.; Dumais, S. & White, T. (2005), 'Evaluating implicit measures to improve web search', *ACM Trans. Inform. Syst.* 23(2), 147-168.
- Frank, A. & Fröhlich, N. & Voit, M. & Zech, A. (2013), 'Tasktyp-Abhängigkeit des Verhältnisses betrachteter zu vorhandenen visuellen Elementen auf Wikipedia-Seiten', Seminararbeit WS 2012/2013, Universtität Regensburg, Lehrstuhl für Informationswissenschaft.
- Gaus, W. (2005), *Dokumentations- und Ordnungslehre*, Springer Verlag, Berlin.
- Giles, J. (2005), 'Internet encyclopedias go head to head', *Nature*(438), pp. 900-901.
- Grosz, B. & Sidner, C. (1986), 'Attention, intention and the structure of discourse', *Computational Linguistics* 12(3), pp. 175 - 204.
- Grzanna, M.; Koch, M. & Riedl, R. (2011), 'Das Gedächtnis der Welt. Zehn Jahre Wikipedia: Ein Experiment im Netz, das vom Engagement seiner Nutzer lebt und die Suche nach dem Wissen verändert hat', *Süddeutsche Zeitung*(8), p. 21.

- Guo, J.; Xu, G.; Li, H. & Cheng, X. (2008), A Unified and Discriminative Model for Query Refinement, in 'Proceedings of Special Interest Group on Information Retrieval 2008', ACM, New York, NY, USA, pp. 379 - 386.
- Guo, Q. & Agichtein, E. (2010), Ready to buy or just browsing?: detecting web searcher goals from interaction data, in 'Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 130-137.
- Guo, Q. & Agichtein, E. (2010), Towards predicting web searcher gaze position from mouse movements, in 'Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems', ACM, New York, NY, USA, pp. 3601-3606.
- Haasebrook, J. (1995), Multimedia-Psychologie: Eine neue Perspektive menschlicher Kommunikation, Spektrum Wissenschaftsverlag, Heidelberg.
- Hall, H. (1981), 'Patterns in the use of information: The right to be different', Journal of the American Society for Information Science 32, pp. 103-112.
- Hammwöhner, R. (2007), 'Qualitätsaspekte der Wikipedia', Wikis: Diskurse, Theorien und Anwendungen. Sonderausgabe von kommunikation@gesellschaft 8.
- Han, S.; Yue, Z. & He, D. (2013), Automatic Detection of Search Tactic in Individual Information Seeking: A Hidden Markov Model Approach, in 'Proceedings of the iConference 2013', pp. 712-716.
- Hanbury, A. (2012), Medical information retrieval: an instance of domain-specific search, in 'Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 1191-1192.

- Harper, D. J. & Kelly, D. (2006), Contextual relevance feedback, in 'Proceedings of the 1st international conference on Information interaction in context', ACM, New York, NY, USA, pp. 129-137.
- Hartley, R. V. L. (1928), 'Transmission on Information', Bell System Technical, pp. 535-563.
- Hassenzahl, M. (2004), 'The Interplay of Beauty, Goodness, and Usability in Interactive Products', Human Computer Interaction 19, pp. 319-349.
- He, Q.; Jiang, D.; Liao, Z.; Hoi, S.; Chang, K.; Lim & Li, H. (2009), Web Query Recommendation via Sequential Query Prediction, in 'Proceedings of the IEEE 25th International Conference on Data Engineering 2009', pp. 1443-1454.
- Head, A. & Eisenberg, M. (2010), 'How today's college students use Wikipedia for course related research', first monday, peer-reviewed journal on the internet 15(3).
- Hearst, M. A. (2009), Search User Interfaces, in R. Baeza-Yates & B. Ribeiro-Neto, ed., 'Information Visualization', Cambridge: University Press.
- Heidenreich, S. (2005), Neue Medien, in K. Sachs-Hombach, ed., 'Bildwissenschaft: Disziplinen, Themen, Methoden', Suhrkamp: Frankfurt/Main, .
- Herczeg, M. (2005), Softwareergonomie, Oldenbourg Verlag.
- Herrschaft, L. (1996): Zur Bestimmung eines medienspezifischen Informationsbegriffs. Nachrichten für Dokumentation 47, 171-182.

- Hu, J.; Wang, G.; Lochovsky, F. & Sun, Jian-Tao und Chen, Z. (2009), Understanding User's Query Intent with Wikipedia, in 'Proceedings of the WWW Conference 2009, Madrid, Spain', pp. 471 -481.
- Huang, J.; White, R. & Buscher, G. (2012), User see, user point: gaze and cursor alignment in web search, in 'Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems', ACM, New York, NY, USA, pp. 1341-1350.
- Huffman, S. & Hochster, M. (2007), How well does result relevance predict session satisfaction?, in 'Proceedings of Special Interest Group on Information Retrieval Conference 2007, pp. 567- 574.
- Ingwersen, P. (1992), 'Information and Information Science', Encyclopedia of Library and Information Science 56(19), pp. 137-174.
- Ingwersen, P. (1982), 'Search procedures in the library analyzed from the cognitive point of view', Journal of Documentation 38, pp. 165-191.
- Ingwersen, P. & Jarvelin, K. (2005), SIGIR Workshop reports: Information Retrieval in Context: IRIx, in 'ACM SIGIR Forum', pp. 31-39.
- Ingwersen, P. & Jarvelin, K. (2005), The turn: Integration of information seeking and retrieval in context, Springer: Dordrecht, NL.
- Izard, C. (2009), 'Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues', The Annual Review of Psychology 60, pp. 1-25.
- James, R. (1983), 'Libraries in the mind: How we can see user's perceptions of libraries', Journal of the Librarianship 15, pp. 19-28.



- Jansen, B. J. & McNeese, M. D. (2005), 'Evaluating the effectiveness of and patterns of interactions with automated searching assistance', *JASIST - Journal of the American Society for Information Science and Technology* 56(14), pp. 1480 - 1503.
- Joachims, T.; Granka, L.; Pan, B.; Hembrooke, H.; Radlinki, F. & Gay, G. (2007), 'Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search', *ACM Trans. Inform. Syst.* 25(2), Article No. 7.
- Jochmann-Mannak, H.; Huibers, T.; Lentz, L. & Sanders, T. (2010), 'Children searching information on the Internet: Performance on children's interfaces compared to Google', *Towards Accessible Search Systems - Workshop of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 27-35.
- Johnson, J. (2002), 'On contexts of information seeking', *Information Processing and Management* 39, pp. 735-760.
- Joos, M.; Rütting, M. & Velichkovsky, B. (2002), 'Bewegungen des menschlichen Auges: Fakten Methoden und innovative Anwendungen', TU Dresden, Institut für Psychologie III, Professur Ingenieurspsychologie und Kognitive Ergonomie.
- Kang, I.-H. & Kim, G. (2003), Query Type Classification for Web Document Retrieval, in 'Proceedings of Special Interest Group on Information Retrieval Conference 2003', ACM Press, Toronto, Canada, pp. 64 - 71.
- Kappas, A. & Müller, M. G. (2006), 'Bild und Emotion - ein neues Forschungsfeld', *Publizistik* 51(1), pp. 3-23.
- Kelly, D. & Belkin, N. J. (2001), Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback, in 'Proceedings of Special Interest Group on Information Retrieval Conference 2001', pp. 408-409.

- Kelly, D. & Cool, C. (2002), The Effects of Topic Familiarity on Information Search Behavior, in 'Proceedings of the Joint Conference on Digital Libraries 2002', ACM Press, pp. 74-75.
- Kelly, D. & Fu, X. (2007), 'Eliciting better information need descriptions from users of information search systems', *Information Processing & Management* 43(1), pp. 30 - 46.
- Kelly, D. & Teevan, J. (2003), Implicit Feedback for Inferring User Preference: A Bibliography. 'SIGIR Forum', AMC Press.
- Kleinginna, Paul, R. & Kleinginna, Anne, M. (1981), 'A categorized list of emotion definitions, with suggestions for a consensual definition', *Motivation and Emotion* 5(4), pp. 345 - 379.
- Knäusl, H.; Elswiler, D. & Ludwig, B. (2012), Towards Detecting Wikipedia Task Contexts, in 'Proceedings of the EuroHCIR '12', ACM Press, Nijmegen, The Netherlands.
- Knäusl, H. & Ludwig, B. (2013), What Readers want to Experience: An Approach to Quantify Conversational Maxims with Preferences for Reading Behaviour, in 'Proceedings of the 5th International Conference on Agents and Artificial Intelligence'.
- Knäusl, H.; Rösch, B. & Schubart, L. (2012), 'Einfluss von Kontextfaktoren auf das Suchverhalten in der Wikipedia', *Information. Wissenschaft und Praxis* 63(5), pp. 319 - 323.
- Knäusl, H. & Schubart, L. (2013), Kontextsensitive Bildauswahl? Studien zur Bild-Präferenz, in 'Proceedings of the Internationales Symposium für Informationswissenschaft 2013', Verlag Werner Hülsbusch, , pp. 349 - 361.

- Kostadinova, H.; Totkov, G. & Indzhov, H. (2012), Adaptive e-learning system based on accumulative digital activities in revised Bloom's taxonomy, in 'Proceedings of the 13th International Conference on Computer Systems and Technologies', ACM, New York, NY, USA, pp. 368-375.
- Kuhlen, R. (2004), Informationsethik, UVK Verlagsgesellschaft: Konstanz.
- Kuhlthau, C. (2004), Seeking meaning: A process approach to library and information services, Libraries: Westport.
- Kuhlthau, C. (1991), 'Inside the search process: Information seeking from the user's perspective', Journal of the American Society for Information Science 42(5), pp. 361-371.
- Kules, B. & Capra, R. (2009), Designing exploratory search tasks for user studies of information seeking support systems, in 'Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries', pp. 419-420.
- Landow, G. P. (2006), Hypertext 3.0: Critical Theory and New Media in an Era of Globalization, The John Hopkins University Press: Baltimore and Maryland.
- Landow, G. P. (1997), Hypertext 2.0: The Convergence of Contemporary Critical Theory and Technology, The John Hopkins University Press: Baltimore and Maryland.
- Lau, R.; Bruza, P. & Song, D. (2008), Towards a Belief-Revision-Based Adaptive and Context-Sensitive Information Retrieval System'ACM Trans. Inf. Syst.', Article 8.
- Lee, U.; Liu, Z. & Cho, J. (2005), Automatic Identification of User Goals in Web Search, in 'Proceedings of the International World Wide Web Conference 2005', pp. 391-400.

- Lenz, W. (1990), *Kleine Geschichte großer Lexika*, Bertelsmann Lexikon Verlag, Gütersloh.
- Leonhart, R. (2013), *Lehrbuch Statistik, Einstieg und Vertiefung, Vol. 3. überarbeitete Ausgabe*, Verlag Hans Huber. Bern.
- Levin, J.; Press, A., ed., (1981), *On Functions of Pictures in Prose*, in: F.J. Pirozzolo; & M.C. Wittrock (eds.) *Neurophysiological and cognitive processes in reading*, New York: Academic Press, pp. 202-228.
- Lewandowski, D. (2005), *Web Information Retrieval, Technologien zur Informationssuche im Internet*, Dinges & Frick GmbH, Wiesbaden, DGI-Schrift (Informationswissenschaft 7).
- Li, W. (2011), *Domain-specific information retrieval using recommenders*, in 'Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval', ACM, New York, NY, USA, pp. 1327-1328.
- Lim, S. (2009), 'How and why do college students use Wikipedia', *Journal of the American Society for Information Science and Technology* 60(11), pp. 2189-2202.
- Lindley, S. E.; Meek, S.; Sellen, A. & Harper, R. (2012), 'It's simply integral to what I do': enquiries into how the web is weaved into everyday life, in 'Proceedings of the 21st international conference on World Wide Web', ACM, New York, NY, USA, pp. 1067-1076.
- Liu, J.; Liu, C.; Cole, M.; Belkin, N. & Zhang, X. (2012), *Exploring and Predicting Search Task Difficulty*, in 'Proceedings of the The 21st ACM International Conference on Information and Knowledge Management', ACM Press, , pp. 1313 - 1322.

- Lommatzsch, A.; Kille, B. & Albayrak, S. (2013), Learning Hybrid Recommender Models for Heterogeneous Semantic Data, in 'Proceedings of the 28th Annual ACM Symposium on Applied Computing', ACM Press, , pp. 275 - 276.
- Lopatovska, I. (2009), Searching for good mood: Examining relationships between search task and mood, in 'Proceedings of the American Society for Information Science and Technology', pp. 1-13.
- Ludwig, B.; Mandl, S.; Elswiler, D. & Mika, S. (2011), REFRESH: Recommendations and feedback for realising an stabilising health, in 'Proceedings of the 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)', pp. 206-207.
- Müller, M. (2003), Grundlagen der visuellen Kommunikation: Theorieansätze und Analysemethoden, UVK Verlagsgesellschaft: Konstanz.
- Machleit, K. A. & Eroglu, S. A. (2000), Describing and Measuring Emotional Response to Shopping Experience', Journal of Buiseness Reaserch 49(2), pp. 101-111.
- Marchionini, G. (1995), Information Seeking in Electronic Environments, Cambridge: University Press.
- Marchionini, G. (1989), 'Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia', Journal of the American Society for Information Science 40(1), pp. 54-66.
- Marchionini, G.; Geisler, G. & Brunk, B. (2000), Agileviews: A Human-Centred Framework for Interfaces to Information SPaces, in 'Proceedings of the Anual Conference ot the American Society for Information Science', pp. 271-280.
- Marchionini, G. & Shneiderman, B. (1988), 'Finding facts vs. browsing knowledge in hypertext systems.', Comuter 21(1), pp.70-80.

- Marchionini, G. (2006), 'Exploratory search: from finding to understanding', *Communications of the ACM* 49(4), pp. 41-46.
- Maslow, A. (1943), *A Theory of Human Motivations*, *Psychological Review* 40(4), pp. 370-396.
- Mayer, R. E. (2001), *Multimedia Learning*, New York: Cambridge University Press.
- McKenzie, P. J. (2003), 'A model of information practices in accounts of everyday-life information seeking', *Journal of Documentation* 59(1), pp. 19-40.
- McNamara, D. (2001), *Speed Reading*. *International Encyclopedia of the Social & Behavioral Science*, Pergamon: Oxford.
- Morita, M. & Shinoda, Y. (1994), Information filtering based on user behavior analysis and best match text retrieval, in 'Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA pp. 272-281.
- Moshfeghi, Y. (2009), Affective adaptive retrieval: study of emotion in adaptive retrieval, in 'Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 852-852.
- Moshfeghi, Y. & Jose, J. (2013), On Cognition, Emotion, and Interaction Aspects of Search Tasks with Different Search Intentions, in 'Proceedings of the International World Wide Web Conference 2013', ACM Press, pp. 931-941.
- Nöth, W. (2008), Der Zusammenhang von Text und Bild, in K. Brinker; G. Antos & W. Heinemann, ed., 'Text- und Gesprächslinguistik', Mouton de Gruyter, pp. 489-496.

- Nahl, D. & Bilal, D. (2007), *Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory*, Information Today, Medford and New Jersey.
- Navalpakkam, V. & Churchill, E. (2012), *Mouse Tracking: Measuring and Predicting User's Experience of Web-based Content*, in 'Proceedings ACM Conference on Human Factors in Computing Systems 2012', pp. 2963 – 2973.
- Nielsen, J. (2006), *Designing Web Usability*, New Riders: Berkeley, California.
- Nielsen, J. (1989), *The matters that really matter for hypertext usability*, in 'Proceedings of the Second ACM Conference on Hypertext', pp. 239-248.
- Ortony, A.; Clore, G. & Collins, A. (1988), *The Cognitive Structure of emotions*, Cambridge University Press: Cambridge.
- Osgood, C. (1952), *'The Nature and Measurement of Meaning'*, in *Psychological Bulletin* 49(3), pp. 172-237.
- Osgood, C. E.; Suci, G. & Tannenbaum, P. (1957), *The Measurement of Meaning*, University of Illinois Press: Illinois.
- Paivio, A. (1971), *Imagery and verbal processes*, New York: Holt, Rinehart and Winston.
- Peez, G. (2006), *Fotoanalyse nach Verfahrensprinzipien der Objektiven Hermeneutik*, in Winfried Marotzki & Horst Niesyto, ed., *'Bildinterpretation und Bilderverstehen'*, VS Verlag für Sozialwissenschaften: Wiesbaden, pp. 121-141.
- Penniman, W. (1975), *'Rhythms of Dialogue in Human-computer Conversation'*, PhD thesis, Ohio State University.

- Picard, R. (1998), 'Towards agents that recognize emotion', Technical report (515), Massachusetts Institute of Technology.
- Poddar, A. & Ruthvan, I. (2010), The Emotional Impact of Search Tasks, in 'Proceedings of the IiiX 2010', ACM Press, pp. 35 - 44.
- Pontes, U. (2011), 'Was sind Emotionen?', Das Gehirn - der Kosmos im Kopf, online verfügbar unter [<http://dasgehirn.info/>]; abgerufen 12.März 2013
- Qiu, L. (1993), 'Markov models of search state patterns in a hypertext information retrieval system', Journal of the American Society for Information Science 44(7), pp. 413-427.
- Radach, R.; Heller, D. & Huestegge, L. (2004), 'Blickbewegungen beim Lesen: Neueste Entwicklungen und Ansatzpunkte für die Legasthenieforschung', in Schulte, G. (ed), Legasthenie: Zum aktuellen Stand der Ursachenforschung, der diagnostischen Methoden und der Förderkonzepte, Verlag Dr. Winkler, Bochum, pp. 61-87.
- Rahulamathavan, Y.; Phan, R.C.-W.; Chambers, J.A. & Parish, D.J. (2012), 'Facial Expression Recognition in the Exnctypted Domain based on Local Fisher Discriminant Analysis', IEEE Transactions on Affektive Computing 4(1), pp. 83 - 92
- Ramos, V. F. & de Bra, P. M. (2010), The Influence of Adaption on Hypertext Structures and Navigation, 'Proceedings of the 21st ACM conference on Hypertext and hypermedia', ACM Press New York, pp. 77 – 82.
- Rayner, K. (1998), 'Eye movements in reading and information processing: 20 years of research', Psych. Bull 124(3), pp. 372-422.



- Reeves, B. & Nass, C. (1998), The Media Equation: How people treat computers, television and new media like real people and places, ELCI Publications 8(1), online verfügbar [\[http://www.stanford.edu/group/cslipublications/cslipublications/site/1575860538.shtml\]](http://www.stanford.edu/group/cslipublications/cslipublications/site/1575860538.shtml) er], abgerufen 21. Oktober 2012
- Reid-Cuninham (2008), 'Malslow's Theory of Motivation and Hierarchy of Human Needs: A Critical Analysis', PhD thesis, School of Social Welfare, University of California - Berkeley.
- Rodden, K. & Fu, X. (2007), Exploring how mouse movements relate to eye movements on web search results pages, in 'SIGIR Workshop on Web Information Seeking and Interaction', pp. 29-32.
- Rodden, K.; Fu, X.; Aula, A. & Spiro, I. (2008), Eye-mouse coordination patterns on web search results pages, in 'Proceedings of the international conference of Human-Computer Interaction CHI '08, extended abstracts on Human factors in computing systems', ACM, New York, NY, USA, pp. 2997-3002.
- Rose, D. & Levinson, D. (2004), 'Understanding User Goals in Web Search', Proceedings of the World Wide Web conference 2004, pp.13-19.
- Russell, J. A. (1979), 'Affective Space is Bipolar', Journal of Research and Personality 37(3), pp. 345-356.
- Russell, J. A. (1980), 'A Circumplex Model of Affect', Journal of Personality and Social Psychology 39(6), pp. 1161-1172.
- Russell, J. A. & Mehrabian, A. (1977), 'Evidence for a Three-Factor Theory of Emotions', Journal of research in Personality 11(3), pp. 273-294.

- Rust, M. (2005), BioVid – ontologiestütztes, kontextsensitives Retrieval biologischer Videosequenzen, 'In proceeding of: INFORMATIK 2004 - Informatik verbindet, Band 2, Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI) ', pp. 54-59.
- Sachs-Hombach, K. (2001), Kann die semiotische Bildtheorie Grundlage einer allgemeinen Bildwissenschaft sein?, in K. Sachs-Hombach, ed., 'Bildhandel: Interdisziplinäre Forschung zur Pragmatik bildhafter Darstellungsformen', Scriptorum Verlag Magdeburg GmbH, pp. 9-26.
- Savolainen, R. (1995), 'Everday life information seeking: Approching information seeking in the context of "way of life"', Library and Information Science Research 17(3), pp. 259-294.
- Schamber, L.; Eisenberg, M. B. & Nilan, M. S. (1990), 'A re-examination of relevance: toward a dynamic, situational definition', Information Processing & Management 26(6), pp. 755 - 776.
- Scherer, K.R., Scherer, K. R., ed., (1990), Theorien und aktuelle Probleme der Emotionspsychologie, Hogrefe, Göttingen.
- Scherer, K. R. & Shorr, A. Johnstone, T., ed. (2001), Appraisal processes in emotion: theory, methods, research, Oxford University Press, Canary, NC.
- Scherer, K. R. (2005), 'What are emotions? And how can they be measured?', Social Science Information 44(4), pp. 695-729.
- Schmauks, D. (2005), Bilder im Kontext von Rätsel und Spiel 'Bildwissenschaften zwischen Reflexion und Anwendung', Herbert von Halem Verlag, Köln, pp. 342-356.

- Schmidt-Atzert, L. (1996), Lehrbuch der Emotionspsychologie, Kohlhammer, Stuttgart.
- Schnell, R.; Hill, P. B. & Esser, E. (2008), Methoden der empirischen Sozialforschung, Oldenbourg Verlag, München Wien.
- Schnotz, W. (2006), Pädagogische Psychologie, Beltz, Weinheim.
- Schubart, L. (2011), 'Bildverwendung in der Wikipedia. Analyse einer automatisierbaren Klassifikation für Bilder.', Bachelorarbeit zur Erlangung des akademischen Grades Bachelor of Arts im Fach Informationswissenschaft, Universität Regensburg, Lehrstuhl für Informationswissenschaft.
- Schubart, L. (2013), 'Einfluss des Bildgegenstandes auf die Bewertung: Untersuchung von Bildern aus den Wikimedia Commons nach bild- und rezeptionsbezogenen Kriterien', Masterarbeit zur Erlangung des akademischen Grades „Master of Arts“ im Fach Informationswissenschaft, Universität Regensburg, Lehrstuhl für Informationswissenschaft.
- Seiffert, H., ed. (1971), Information über die Information: Verständigung im Alltag - Nachrichtentechnik - Wissenschaftliches Verstehen - Informationssoziologie. Das Wissen des Gelehrten, Beck, München.
- Shen, D.; Sun, J.-T.; Yang, C. & Zheng, C. (2006), Building bridges for web query classification, in 'Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval', ACM, New York, NY, USA, pp. 131-138.
- Shen, X. (2007), 'User-centered adaptive information retrieval', Phd Thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA.

- Shen, X.; Tan, B. & Zhai, C. (2005), Context-Sensitive Information Retrieval Using Implicit Feedback, in 'Proceedings of Special Interest Group on Information Retrieval 2005', pp. 43 – 50.
- Simons, J. (1999), 'What's a digital image?', in: Bild-Medium-Kunst, Spielmann, Y. & Winter, G. (eds), Wilhelm Fink Verlag, pp. 107 - 122.
- Singh, R.; Hsu, Y.-W. & Moon, N. (2013), 'Multiple perspective interactive search: a paradigm for exploratory search and information retrieval on the web', *Multi-media Tools Appl.* 62(2), pp. 507 - 543.
- Smith, J. & Chang, S. (1997), 'Multi-Stage Classification of Images from Features and Related Text', *Proceedings Fourth DELOS workshop*.
- Sohn, T.; Li, K. A.; Griswold, W. G. & Hollan, J. D. (2008), A Diary Study of Mobile Informations Needs', *Proceedings of the Special Interest Group on Computer Human Interaction Conference 2008*, pp. 433-442.
- Somol, P.; Pudil, P.; Novovicova, J. & Paclik, P. (1999), 'Adaptive floating search methods in feature selection', *Pattern Recognition Letters*(20), pp. 1157-1163.
- de Sousa, R. (1990), *The rationality of emotions*, Cambridge: MIT Press.
- Spink, A. & Cole, C. (2001), 'Introduction to the special issue: Everyday life information-seeking research', *Library & Information Science Research* 23, pp. 301-304.
- Spoerri, A. (2007), 'What is popular on Wikipedia and why?', first monday, peer-reviewed journal on the internet 12(4), online verfügbar unter [<http://ojs-prod-lib.cc.uic.edu/ojs/index.php/fm/article/view/1765/1645>].
- Stegbauer, C. (2012), Dossier Wikipidia: 'Die Macht der Wenigen, Soziale Prozesse und Strukturen', Bundeszentrale für politische Bildung, 10.10.2012, online verfügbar

unter [http://www.bpb.de/gesellschaft/medien/wikipedia/145809/die-macht-der-wenigen?p=all].

Steinmüller, W. (1993), *Informationstechnologie und Gesellschaft: Einführung in die angewandte Informatik*, Wissenschaftliche Buchgesellschaft, Darmstadt.

Strong, D. M.; Lee, Y. W.; Wang, R. Y. & Strong, D. M. (1997), 'Data Quality in Context', *Communications of the ACM* 40(5), pp. 103 - 110.

Stvilia, B.; Twidale, M. B.; Gasser, L. & Smith, L. (2005a), 'Information Quality Discussions in Wikipedia', *Knowledge Management: Nurturing Culture, Innovation, and Technology - Proceedings of the 2005 International Conference on Knowledge*, pp. 101-113.

Stvilia, B.; Twidale, M. B.; Smith, L. C. & Gasser, L. (2005b), 'Assessing information quality of a community-based encyclopedia', *Proceedings of the International Conference on Information Quality - ICIQ 2005*, pp. 442-454.

Swanson, T. (2004), 'A radical step: Implementing a critical information literacy model', *Portal. Libraries and the Academy* 4(2), pp. 259-273.

Taylor, R. (1962), 'The process of asking questions', *Journal of the American Society for Information Science and Technology* 13(4), pp. 391-396.

Tenenbaum, J. (2000), 'Rules and Similarity in Concept Learning', *Advances in Neural Information Processing Systems* 12, pp. 59 - 65.

Trumbley, J.; Arnett, K. & Johnson, P. (1994), 'Productivity Gains via an Adaptive User Interface', *Journal of human-computer studies* 40, pp. 63-81.

Viegas, F. B. (2006), 'The Visual Side of Wikipedia', *Proceedings of 40th Annual Hawaii International Conference on Systems Sciences*, p. 85.

- Vogel, J. (2013), 'Bild oder Text? Präferenzen der Nutzer für verschiedene Inhaltstypen in Abhängigkeit der durchsuchten Domain', Masterarbeit zur Erlangung des akademischen Grades „Master of Arts“ im Fach Informationswissenschaft, Universität Regensburg, Lehrstuhl für Informationswissenschaft.
- Vyzas, E. (1999), 'Recognition of emotional and cognitive states using physiological data', Technical report (510), Massachusetts Institute of Technology, Boston.
- Wang, P. & Soergel, D. (1998), 'A cognitive model of document use during a reasearch projekt', Journal of the American Society for Information Science 49(2), pp. 115-133.
- Wang, R. & Strong, D. (1996), 'Beyond Accuracy: What Data Quality Means to Data Consumers', Journal of Management Information Systems 12(4), pp. 5-33.
- Wang, R. Y.; Storey, V. C.; Firth, C. P. & Wang, R. Y. (1995), 'A Framework for Analysis of Data Quality Research', IEEE Transactions on knowledge and data engineering 7(4), pp. 623-640.
- Weidenmann, B. (1991), Lernen mit Bildmedien: Psychologische und didaktische Grundlagen, Beltz, Weinheim.
- Wen, Z.; Zhou, M. X. & Aggarwal, V. (2007), Context-Aware, adaptive information retrieval for investigative tasks, in 'Proceedings of the 12th international conference on Intelligent user interfaces', ACM, New York, NY, USA, pp. 122-131.
- Wersig, G. (1971), Information-Kommunikation - Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft, Verl. Dokumentation, München.

- White, R. W.; Dumais, S. & Teevan, J. (2009), Characterizing the Influence of Domain Expertise on Web Search Behavior, in 'Proceedings of the second ACM International Conference on Web Search and Datamining', ACM Press, pp. 132-141 .
- White, R. W. & Kelly, D. (2006), A study on the effects of personalization and task information on implicit feedback performance, in 'Proceedings of the 15th ACM international conference on Information and knowledge management', ACM, New York, NY, USA, pp. 297-306.
- White, R. W.; Ruthven, I. & Jose, M. J. (2005), 'A Study of Factors Affecting the Utility of Implicit Relevance Feedback', Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 35-42.
- Wiegand, D. (2007), 'Entdeckungsreise, Digitale Enzyklopädien erklären die Welt', c't, Magazin für Computer und Technik(6), pp. 136-145.
- Wikipedia (2010), 'Kategorien', online verfügbar unter [\[http://de.wikipedia.org/wiki/Wikipedia:Kategorien\]](http://de.wikipedia.org/wiki/Wikipedia:Kategorien), abgerufen September 2010.
- Wikipedia (2011a), 'Themenportale', online verfügbar unter [\[http://de.wikipedia.org/wiki/Portal:Wikipedia\\_nach\\_Themen\]](http://de.wikipedia.org/wiki/Portal:Wikipedia_nach_Themen), abgerufen Juni 2011.
- Wikipedia (2011b), 'Kategorie:~Inselgruppe als Thema', online verfügbar unter [\[http://de.wikipedia.org/wiki/Kategorie:Inselgruppe\\_als\\_Thema\]](http://de.wikipedia.org/wiki/Kategorie:Inselgruppe_als_Thema), abgerufen Juni 2011
- Wikipedia (2011c), 'Kategorie: Räumliche Systematik', online verfügbar unter [\[http://de.wikipedia.org/wiki/Kategorie:R%C3%A4umliche\\_Sachsystematik\]](http://de.wikipedia.org/wiki/Kategorie:R%C3%A4umliche_Sachsystematik), abgerufen Juli 2011

- Wildemuth, B. (2004), 'The Effects of domain knowledge on search tactic formulation', *Journal of the American Society for Information Science and Technology* 55(3), pp. 246-258.
- Wildemuth, B. & Freund, L. (2009), Search Tasks and Their Role in Studies of Search Behavior, in 'Proceedings of the Third anual workshop of Human Computer Interaction 2009'.
- Wildemuth, B.; Freund, L. & Toms, E. (2010), 'Systematic Review of Imposed Search Tasks', online verfügbar unter [<http://ils.unc.edu/searchtasks/>], abgerufen September 2013
- Willrich, R.; de Moura Speroni, R.; Lima, C. V.; de Oliveira Diaz, A. L. & Penedo, S. M. (2006), Adaptive information retrieval system applied to digital libraries, in 'Proceedings of the 12th Brazilian Symposium on Multimedia and the web', ACM, New York, NY, USA, pp. 165-173.
- Wilson, T. (2006), 'On user studies and information needs', *Journal of Documentation* 62(6), pp. 658-670.
- Wilson, T. (1977), *Public Knowledge, Private Ignorance*, Greenwood, Westport, CT.
- Wilson, T. D. (2000), 'Human Information Behaviour', *Information Science* 3(2), pp. 49-55.
- Wilson, T. D. (1999), 'Models in information behavior research', *Journal of Documentation* 55(3), pp. 249-270.
- Wissenmedia GmbH (2009-2010), 'Brockhaus Enzyklopädie', Gütersloh.



- Wittgenstein, L. (1953), *Philosophische Untersuchungen - Philosophical investigations*, Oxford.
- Wittman, M. (2012), 'Emotionen bei der Suche in der Wikipedia', Masterarbeit zur Erlangung des Grades „Master of Arts“ im Fach Informationswissenschaft, Universität Regensburg, Lehrstuhl für Informationswissenschaft
- Wu, I.-C. & Wu, C.-Y. (2011), 'Using internal link and social network analysis to support searches in Wikipedia: A model and its evaluation', *Journal of the Information Science* 37(2), pp. 189-207.
- Wundt, W. (1896), *Grundriss der Psychologie*, Entgelmann, Leibzig.
- Yan, X.; Lau, R. Y.; Song, D.; Li, X. & Ma, J. (2011), 'Toward a semantic granularity model for domain-specific information retrieval', *ACM Trans. Inf. Syst.* 29(3), pp. 15:1-15:46.
- Yue, Z.; Han, S. & He, D. (2013), *Automatic Detection of Search Tactics in Collaborative Exploratory Web Search Prozess*, CSCW Workshop on Collaborative Information Seeking (CSCW 2013). Feb. 23-27, 2013, San Antonio, Texas, USA.
- Zhang, X.; Asano, Y. & Yoshikawa, M. (2011), 'Towards Improving Wikipedia as an Image-Rich Encyclopaedia through Analyzing Appropriateness of Images for an Article', *Proceedings of the 13th Asia-Pacific web conference on Web technologies and applications*, pp. 200 – 212.
- Zini, F. & Ricci, F. (2012), *Guiding patients in the hospital*, in 'Advances in User Modeling, Proceedings of the The 19st Conference on User Modeling, Adaptation and Personalization UMAP 2011', Springer Verlag Berlin Heidelberg, pp. 309-319.



## 20 Anhang

### 1. Klassifikation für Bilder mit [Schubart, 2012]

Vollständige Klassen

**Semantische Klassen:**

Bildgegenstand	Person/Personen	Lebewesen	Szene	Landschaft/Ort	Gebäude	Gegenstand	Piktogramm
Äquivalenzklassen	Porträt	Flora	Gruppenbild	Region/Land	Architektur	(lebloses) Ding	Flagge
	Personengruppe	Mensch/ Menschenteile		Stadt			
	Person	Fauna	Geschichtlicher Ausschnitt	Ortschaft		Sachverhalt	Wappen
		Mikroorganismen und Bakterien	Momentaufnahme	Schauplatz	Stadt		Insigne
		Erfundene Lebewesen	Zeitpunkte	Gewässer	Innenraum		Karte
		Medizinische Aufnahme	Stillleben	Gebirge/Berg			Symbol
		Menschenmenge (=anonyme Personen)		Sehenswürdigkeit			

Gegenstandsbereich	Kunst und Kultur	Technik	Architektur	Wirtschaft und Politik	Sport	Wissenschaft	Gesellschaft und Alltag	Geographie
Äquivalenzklassen	Mode		Gebäude	Gewerbe und Handel	Sportarten	Tierkunde	Völkerkunde	Orte
	Museen	Maschinen, Geräte, Apparate	Bauwerke	Industrie	Organisationen	Botanik	Soziale Organisationen	Länder
	Theater	Verkehrswesen	Sehenswürdigkeiten	Geld und Währung	Sportveranstaltungen	Medizin	Feste und Bäume	Kontinente
	Film			Staatliche Organisationen	Sportpreise	Geisteswissenschaften	Haushalt	Regionen

	Bilden de Kunst			Handwerk		Naturwissensch haft	Freizeit	
	Literat ur			Tagespolitik/histor ische Politik			Gesundheit	
	Musik			Recht			Fernsehen	

**Syntaktische Klassen:**

Medien- typ	Animation	Bilder	Videos	Computergraphik
	Animiertes Bild	Foto, Scan, Medi- zinische Aufnah- men...	Bewegtes Bild	Kein Abbildungs- prozess, im Com- puter entstanden

Bildgröße	Klein	Unteres mittel	Mittleres mittel	Großes Mittel	Groß
-----------	-------	----------------	---------------------	---------------	------

Bildfarbe	Farbig	Nicht farbig
-----------	--------	--------------

Position in Bezug auf den Paragraphen	Anfang links	Anfang rechts	Mitte links	Mitte rechts	Ende links
---	--------------	---------------	-------------	--------------	------------

Benachbarte Bilder	Erstes von Vielen (viele >10)	Eines von vielen	Erstes von wenigen	Eines von wenigen	Einziges Bild
-----------------------	----------------------------------	------------------	-----------------------	-------------------	---------------

**Pragmatische Klassen:**

<b>Bezug zum Artikelgegenstand</b>	Vollständiger Bezug zum Artikelgegenstand	Bezug zu Teilen des Artikelgegenstand	Sehr loser Bezug zum Artikelgegenstand	Kein Bezug zum Artikelgegenstand
	100	66	33	0

<b>Bezug zum Textabschnitt</b>	Vollständiger Bezug zum Textabschnittes	Bezug zu Teilen des Textabschnittes	Sehr loser Bezug zum Textabschnitt	Kein Bezug zum Textabschnitt
	100	66	33	0

<b>Verwendungsart</b>	Ergänzend informativ	Ergänzend überflüssig	Dekorativ-repräsentierend	Veranschaulichend-repräsentierend
-----------------------	----------------------	-----------------------	---------------------------	-----------------------------------

<b>Bild und Bilduntertitel</b>	Bild verständlich ohne Untertitel	Bild besser verständlich mit Untertitel	Bild schwer verständlich ohne Untertitel	Bild nicht verständlich ohne Untertitel
	100	66	33	0

## 2. Ergebnisse der Mozartstudie

Test-person 2	Test-person 3	Test-person 4	Test-person 5	Test-person 6	Test-person 7	Test-person 8	Test-person 9	Test-person 10	Test-person 11	Test-person 12	Test-person 13	Test-person 14	Test-person 15	Test-person 16
1	1	4	1	1	2	1	3	2	1	1	1	1	1	2
3	5	4	1	1	2	3	4		3	2	4	1	4	3
4	2	2	1	3	3	2	3	3	4	3	3	3	2	1
5	2	4	1	3	4	3	3	2	2	3	2	2	4	3
1		3	1	4	2	1	4	4	1	1	3	1	1	1
2		1	1;3;4;5;7;9;	1;5	1;4;6;8	8	5;7	1	5	4;6;8	4;5;6;9;10	1;2	1;5	7
3		2	2	2	2	2	5	3	1	2	2	1	2	1
2		3	3	1	1	3	2	2	2	4	4	2	3	4
2		1	2	1	3	2	2	2	2	3	2	1	1	1
4		3	4	3	2	2	3	4	3	2	4	1	1	2
2		4	1	1	1	4	3	1	3	keines	3	3	3	1
1;5;7		7	3	4	3	7	2;6	1;5	4		1;10	7	1;2;5;7	4
2		4	1	2	2	2	3	2	1	3	2	1	1	2
4		4	4	1	2	2	4	2	3	3	1	1	5	1
2		2	2	1	4	2	3	3	1	2	3	1	1	3
6		4	3	2	4	2	3	4	3	3	5	3	2	4
3		3	3	3	2	3	3	3	3	3	2	3	3	3
1;5;7		1	1;5;7	1;4	4;7	7	6;8	6	7	2;6;8	1;2;6	1;4	1;2;4;5;6	4
2		2	2	1	2	2	1	1	1	1	5	1	2	1
3		1	4	1	1	1	2	1	2	2	4	1	2	1
5		3	3	1	1	2	2	1	1	2	3	3	3	2
3		3	2	2	3	2	3	3	1	2	1	1	2	2
1		3	1	3	4	4	keines	4	1	1	4	4	4	1
8		6	1;5;6	2;6	1;4;5;6;7	8		1;6	1	2;7;8	8	2;5;7	1;5;6	6
2		3	1	3	3	4	3	3	2	3	5	2	3	3
4		4	2	2	2	2	2	4	3	3	3	2	4	1
3		3	2	3	3	3	2	2	3	2	2	2	3	1
5		4	3	1	4	3	3	3	2	2	4	3	2	3
1		3	1	1	3	2	3	4	4	3	keines Thema des Textes braucht keine Illustration, kein Bild passt	1	4	3
8		1	1;5;7	5;6	6;8	6	6;7	8	1;6	4;7;8		1;5	1;2;5	9

### 3. Tasks aus Eyetrackerstudie eins:

*Tabelle 20-1 Aufgaben Studie eins SS 2012*

#### **Lookup:**

**L1)** Sie sehen zum hundertsten Jahrestag des Titanik Unglücks eine Reportage. Als Sie nach einer Werbepause zu spät zurückschalten, entgeht Ihnen die Information über die genaue Anzahl der Passagiere. Sie beschließen diese Information in der Wikipedia zu recherchieren.

**L2)** Sie diskutieren mit Ihrem Freund über die Oskarverleihung. Nach einiger Zeit geraten Sie in einen Disput darüber, ob Avatar oder Herr der Ringe - Die Gefährten des Rings mehr Oskars erhalten hat. Nach einem heftigen Wortgefecht beschließen Sie die Wikipedia aufzusuchen.

---

#### **Learn:**

**LE1)** Du kommst ursprünglich aus Berlin und studierst seit einer Woche in Regensburg. Am Wochenende haben sich deine Eltern angekündigt. Du möchtest Ihnen eine kleine Stadtführung geben und willst dich deswegen ein bisschen über die Geschichte Regensburgs in der Wikipedia informieren.

**LE2)** Du hast heute Abend ein Date mit einer/m Lacrosse SpielerIn. Ihr wollt euch gemeinsam in einer Sportbar ein Spiel anschauen. Um genügend Gesprächsstoff für den Abend zu haben und weil du nicht als unwissend dastehen möchtest, informierst du dich über die grundlegenden Spielregeln und die wichtigsten Informationen zum Spiel. Die Wikipedia hat sich hier schon öfter als Nachschlagewerk profiliert.

---

#### **Browse:**

**B1)** Man teilt dem Probanden mit, er befindet sich jetzt an einem Rechner im CIP Pool und hat bis seine Veranstaltung beginnt, noch ein paar Minuten Zeit, die er mit dem freien Surfen auf Wikipedia verbringen kann.

**B2)** Die Probanden werden dazu aufgefordert sich nach Themengebieten in der Wikipedia umzusehen. Diese Themengebiete sind letztlich frei wählbar und können sich nach Interessen oder Hobbies der Probanden sowie anderen von uns angegebenen Gebieten richten. Innerhalb des Themengebiets kann sich der Proband allerdings völlig frei bewegen.

#### 4. Fragebögen der Studie eins im SS 2012:

##### 4.1. Pre Study (Testperson#      )

---

1. Alter:

2. Geschlecht:      weiblich    O      männlich    O

3. Studium:      ja      O      nein      O

3a) Falls ja: Studienfach/-

fächer: \_\_\_\_\_

3b) Falls ja: Hochschulsemester: \_\_\_\_\_ BA/MA (bitte nicht zutreffendes streichen)

4. Berufstätig: :      ja      O      nein      O

4a) Falls ja : \_\_\_\_\_

---

5. Wie schätzen Sie ihre Erfahrung ein in Bezug auf Internet/Wikipedia Recherchen?

unerfahren

0      1      2      3      4      5

sehr erfahren

---

6. How strongly do you agree with the following statements?

I use wikipedia as an informational tool

0      1      2      3      4      5

I use wikipedia for my work

0      1      2      3      4      5

0      1      2      3      4      5

---



---

I use wikipedia as a form of entertainment

0 1 2 3 4 5

How long does a typical wikipedia session for you last?

0 1 2 3 4 5

What would you say is the longest you will stay  
on wikipedia for?

0 1 2 3 4 5

---

4.2. Pre Task (Task#            ), (Testperson#            )

1. Wie interessant finden Sie das Thema?

Völlig uninteressant      0      1      2      3      4      5      Sehr interessant

2. Wie gut kennen Sie sich bei diesem Thema aus?

Überhaupt	nicht	0	1	2	3	4	5	Sehr gut
-----------	-------	---	---	---	---	---	---	----------

3. Wie hoch schätzen Sie den Aufwand zur Bearbeitung?

Gering 
0
1
2
3
4
5
 Hoch

4. Wie realistisch finden Sie die Aufgabenstellung?

Völlig unrealistisch      0      1      2      3      4      5      sehr realistisch

## 4.3. Post Task (Task # ), Testperson (# )

## 5. Wie beurteilen sie den Suchverlauf?

Schlecht 0 1 2 3 4 5 zielgerichtet, gut

6. Wie beurteilen Sie ihren Erfolg bei der Lösung der Aufgabe (subjektiv<sup>73</sup>)?

nicht erfolgreich 0 1 2 3 4 5 Sehr erfolgreich

## 7. Wie hoch war der Aufwand zur Bearbeitung?

Gering 0 1 2 3 4 5 Hoch

## 8. Wie fühlen Sie sich? Wählen Sie die passenden Männchen aus!

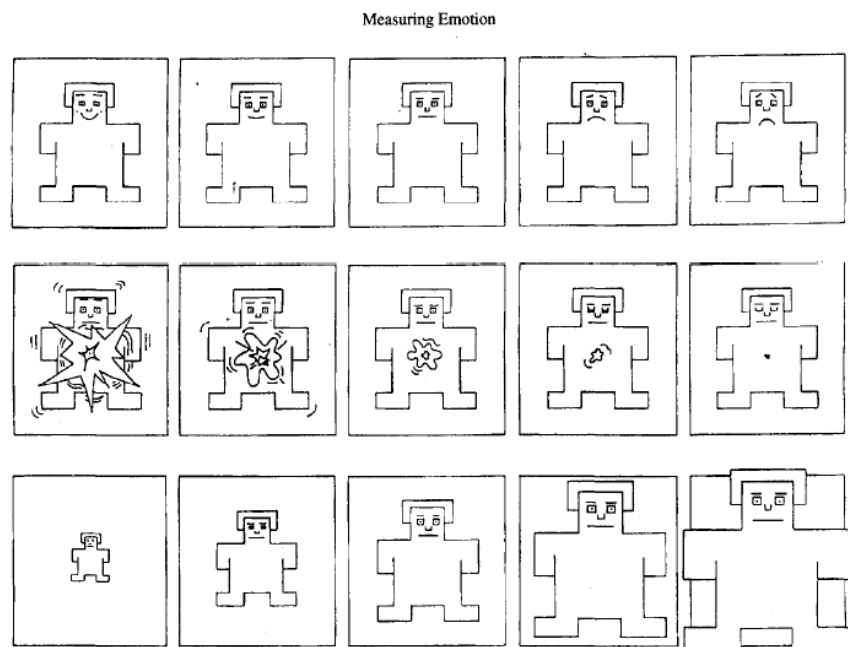


Figure 1. The Self-Assessment Manikin (SAM) used to rate the affective dimensions of valence (top panel), arousal (middle panel), and dominance (bottom panel).

<sup>73</sup> Beim Browsing ist die „Aufgabe“, sich frei zu informieren über Themen, die die Testperson interessieren. Erfolg sollte in diesem Fall daran gemessen werden, ob die Zeit als angenehm und informativ empfunden wurde.

#### 4.4. Post Study (Testperson# )

1. Wie gut geeignet fanden Sie die Wikipedia, um die vorgegebenen Aufgaben zu erfüllen?

gar nicht geeignet 
0
1
2
3
4
5
 sehr gut geeignet

2. Könnten Sie sich eine Quelle vorstellen, die besser geeignet wäre? Wenn ja, welche?

---

---

---

3. Ist Ihnen bei der Bearbeitung der Aufgaben etwas aufgefallen, dass Sie als positiv oder hilfreich wahrgenommen haben (strukturell und/oder inhaltlich)?

---

---

---

4. Ist Ihnen bei der Bearbeitung der Aufgaben etwas aufgefallen, dass Sie als negativ oder störend wahrgenommen haben (strukturell und/oder inhaltlich)?

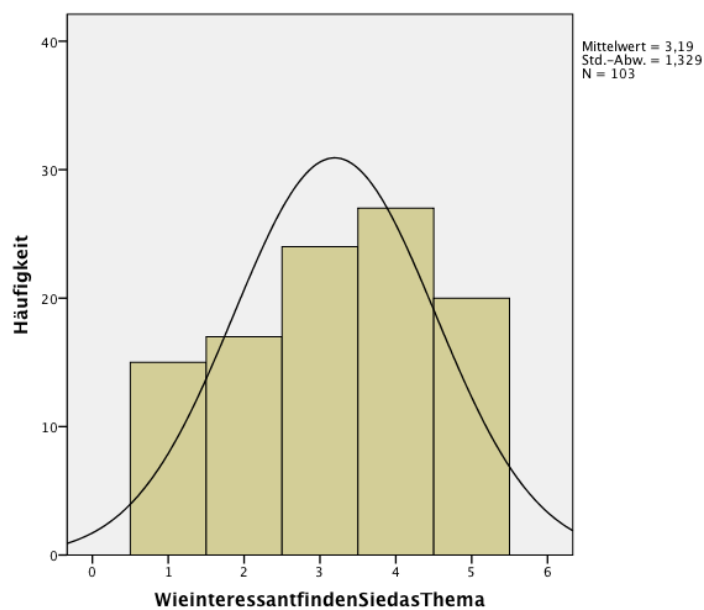
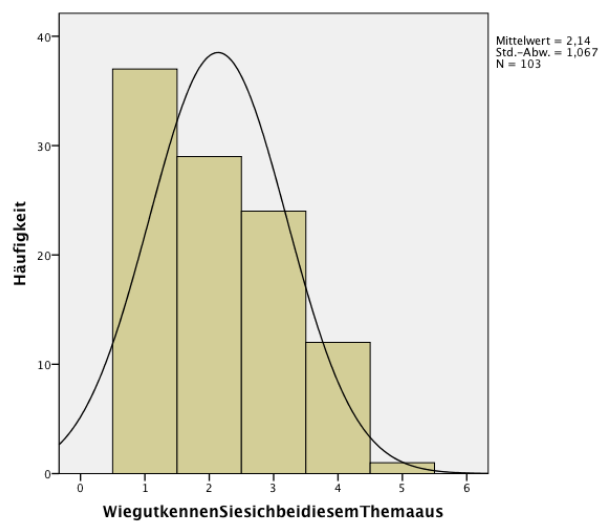
---

## 5. Daten aus Studie eins

Für die statistische Analyse wurden so weit nicht anders angegeben jeweils Microsoft Office Excel 2007 und 54 IBM SPSS Statistics 21 verwendet

### 5.1. Verteilung von Interesse und Themenvorkenntnis, alle Aufgaben außer B

Deskriptive Statistik				
	N	Mittelwert	Standardabweichung	Varianz
WieinteressantfindenSiedasThema	103	3,19	1,329	1,766
WiegutkennenSiesichbeidiesemThema aus	103	2,14	1,067	1,138
Gültige Werte (Listenweise)	103			



## 5.2. Beispiel Videolabels aus Anvil, extrahiert

Tabelle 20-2 Videolabels extrahiert P02-screcc-annotiert-LU (Look-up Task) Studie 1

myID	start	end	duration	Aktionstyp	Infoelementtyp	Navelementtyp
0	4.77	4.77				
4.77	10.7	5.93	NA			
10.7	16.37	5.67	RE			
16.37	16.83	0.47	NA			
16.83	19.5	2.67	RE			
19.5	20.3	0.8	NA			
20.3	21.07	0.77	EX			
21.07	23.9	2.83	RE			
23.9	29.73	5.83	SC			
29.73	37.73	8	RE			
37.73	45.9	8.17	RE			
45.9	46.87	0.97	SC			
46.87	52.13	5.27	RE			
52.13	58.93	6.8	SC			
58.93	62.07	3.13	SC			
62.07	68.13	6.07	RE			
68.13	69.5	1.37	SC			
69.5	71.17	1.67	RE			
71.17	75.13	3.97	SC			
75.13	76	0.87	SC			
76	77	1	SC			
77	79	2	RE			
79	84.27	5.27	SC			
84.27	85.67	1.4	SC			
85.67	86.33	0.67	SC			
86.33	89.5	3.17	SC			
89.5	94.77	5.27	RE			
94.77	97.2	2.43	EX			
97.2	107.87	10.67	RE			
107.87	108.87	1	RE			
10.7	16.37	5.67		LI		
16.83	19.5	2.67		LI		
20.3	21.07	0.77		IG		
21.07	23.9	2.83		IN		
23.9	29.73	5.83		IX		
29.73	37.73	8		IB		
37.73	45.9	8.17		IB		
45.9	46.87	0.97		TX		
46.87	52.13	5.27		IX		
52.13	58.93	6.8		TX		
58.93	62.07	3.13		HL		
62.07	68.13	6.07		IX		
68.13	69.5	1.37		LI		

## 5.3. Häufigkeiten und Lageparameter absolute Häufigkeiten Studie eins

Tabelle 20-3 Absolute Häufigkeiten alle Tasks Studie eins

	BE- BI	BE- IB	BE- IG	LE- AA	LE- EL	LE- IB	LE- LI	LE- LK	LE- UES	QL- AA	QL- EL	QL- IB	QL- IG	QL- LI	QL- LK	QL- UES
VP5 B2	45	0	200	138	3	0	0	2	8	46	47	0	0	57	0	64
VP5 B1	0	0	0	17	10	0	0	0	0	96	7	0	0	45	113	41
VP2 B1	3	0	28	428	37	0	27	0	0	33	5	0	0	11	0	17
VP2 B2	3	0	28	428	37	0	27	0	0	33	5	0	0	11	0	17
VP11 B1	107	0	51	195	86	0	25	3	7	81	2	3	0	35	17	27
Vp11 B2	107	0	51	195	86	0	25	3	7	81	2	3	0	35	17	27
VP6 B1	18	0	9	100	0	5	5	0	14	221	33	38	19	86	0	29
VP6 B2	18	0	9	100	0	5	5	0	14	221	33	38	19	86	0	29
VP10 B2	38	0	13	309	23	0	19	22	3	63	3	0	0	48	24	65
VP10 B1	18	0	88	131	101	8	5	40	27	26	4	24	0	27	32	85
VP9 B2	31	11	0	175	63	0	22	0	0	219	20	0	0	67	0	54
VP9 B1	8	0	0	492	40	0	22	0	0	34	0	8	0	6	0	66
VP8 B2	0	0	0	80	80	52	0	0	15	72	0	0	0	14	0	3
VP8 B1	0	0	0	64	0	0	0	0	0	8	0	0	0	0	0	0
Vp7 B2	75	0	136	154	19	0	6	0	4	54	14	0	0	79	3	113
VP7 B1	37	0	302	32	120	0	42	0	5	15	3	0	0	11	0	47
VP19 B2	14	0	61	386	17	0	0	0	0	13	0	16	0	14	18	70
VP19 B1	13	0	63	208	100	0	115	15	0	10	0	0	0	37	48	0
VP18 B1	38	0	10	0	123	0	0	0	10	168	27	2	0	155	28	48
VP18 B1	38	0	10	0	123	0	0	0	10	168	27	2	0	155	28	48
VP17 B1	32	0	0	152	68	0	140	39	1	12	0	0	0	75	76	4
VP17 B2	129	0	16	189	37	0	76	8	0	2	4	8	0	36	51	9
VP16 B2	20	0	34	388	30	0	50	0	1	5	0	0	0	41	0	102
VP16 B1	29	0	0	585	0	0	0	0	0	0	8	0	0	22	12	0
VP15 B2	0	0	0	278	43	0	78	0	1	48	40	12	0	111	0	43
VP15 B2	0	0	0	278	43	0	78	0	1	48	40	12	0	111	0	43
VP14 B2	93	6	7	102	0	27	0	0	1	159	1	5	0	39	12	226
VP14 B1	97	3	51	0	0	0	0	0	0	61	15	4	0	238	62	120
VP12 B2	85	0	0	523	51	33	0	0	0	0	0	0	0	10	0	2
VP12 B2	85	0	0	523	51	33	0	0	0	0	0	0	0	10	0	2
VP8 LE2	31	0	5	316	147	48	0	0	65	20	0	0	0	4	0	0
VP8 LE1	23	0	29	340	0	0	12	0	9	121	0	0	0	41	0	0
VP7 LE2	82	0	0	461	0	0	7	0	0	31	0	0	0	13	0	101
VP7 LE1	53	0	0	184	64	62	0	0	13	107	19	0	0	94	5	46
VP6 LE2	31	0	0	206	25	0	0	0	6	42	0	0	0	8	0	24
VP6 LE1	5	0	0	138	6	0	0	0	2	344	18	0	0	47	0	29
VP3 LE1	18	0	46	478	14	0	58	0	0	22	0	0	0	0	0	57
VP3 LE2	18	0	46	478	14	0	58	0	0	22	0	0	0	0	0	57
VP2 LE1	32	0	32	397	0	0	37	0	0	39	1	0	0	39	0	60
VP2 LE2	32	0	32	397	0	0	37	0	0	39	1	0	0	39	0	60

VP9 LE1	47	2	0	211	146	0	0	0	0	55	4	0	0	78	11	69
VP9 LE2	75	0	0	285	38	0	36	0	0	119	3	0	0	9	46	59
VP10 LE2	35	0	0	552	38	0	1	0	17	39	0	0	0	1	0	8
VP10 LE1	59	0	4	189	51	0	2	12	25	105	12	1	0	60	0	138
VP11 LE1	129	0	1	96	132	6	28	0	6	176	1	2	0	0	0	80
VP11 LE2	129	0	1	96	132	6	28	0	6	176	1	2	0	0	0	80
V19 L2	41	0	0	626	0	0	0	0	0	4	0	0	0	4	0	14
V19 LE1	52	0	0	371	29	0	32	0	10	98	4	0	0	9	35	3
V18 LE2	10	0	0	286	33	0	0	0	0	44	22	0	0	47	6	158
V18 LE1	10	0	0	286	33	0	0	0	0	44	22	0	0	47	6	158
V17 LE 2	81	0	0	524	47	0	0	0	0	0	0	0	0	1	0	31
V17 LE1	52	0	1	349	13	0	7	0	9	186	4	11	0	6	15	17
V16 LE2	16	0	0	539	42	0	0	0	0	22	0	0	0	17	0	54
V16 LE1	31	0	0	511	52	0	0	0	0	34	0	0	0	0	8	48
V15 LE2	46	0	0	293	6	0	66	0	5	78	0	0	0	37	0	169
V15 LE2	126	0	36	315	58	0	1	0	0	63	0	0	0	16	0	40
VP14 LE1	83	0	0	259	29	0	0	0	0	75	1	0	0	73	0	165
V14 LE1	109	0	12	178	0	0	0	0	0	122	18	5	0	25	0	227
VP12 LE2	53	0	0	277	0	0	179	0	0	45	3	0	0	16	0	28
VP12 LE2	53	0	0	277	0	0	179	0	0	45	3	0	0	16	0	28
VP11 L2	1	0	0	11	0	0	18	0	0	26	16	6	0	28	15	1
VP11 L2	0	0	0	0	0	47	0	0	2	0	0	5	0	19	10	0
VP10 L2	0	0	0	43	0	0	23	20	2	0	5	9	0	48	0	0
VP10 L1	1	0	38	11	0	4	0	0	6	134	5	19	0	36	0	0
VP9 L2	1	0	0	52	5	3	5	0	0	34	4	0	0	78	0	0
VP9 L1	3	0	0	0	0	106	0	0	0	4	19	13	0	67	36	28
VP8 L2	5	0	0	0	0	0	16	0	0	0	4	22	0	144	0	105
VP8 L1	0	0	0	0	0	51	100	0	0	0	33	0	0	0	0	0
VP7 L2	1	0	0	26	0	17	0	0	2	74	125	69	0	66	3	77
VP7 L1	7	7	5	0	0	54	19	0	0	73	0	0	0	5	7	10
VP6 L1	0	0	0	0	0	0	32	0	2	31	12	0	0	40	0	9
VP6 L2	0	0	0	0	0	0	32	0	2	31	12	0	0	40	0	9
VP5 L1	0	0	0	0	0	0	0	0	0	111	3	0	0	27	0	0
VP5 L2	0	0	0	0	0	0	0	0	0	111	3	0	0	27	0	0
VP3 L2	0	0	42	0	0	0	6	0	0	41	0	0	0	0	0	0
VP3 L1	0	0	42	0	0	0	6	0	0	41	0	0	0	0	0	0
VP19 L2	0	0	0	72	0	0	25	0	0	0	0	0	0	13	53	25
VP19 L1	1	0	0	55	0	0	0	63	0	0	2	15	0	97	0	0
VP18 L2	3	0	0	45	0	0	0	0	0	59	3	14	0	67	0	85
VP18 L1	3	0	0	45	0	0	0	0	0	59	3	14	0	67	0	85
VP17 L2	2	0	1	0	30	0	37	0	1	68	3	4	0	52	9	32
VP17 L1	0	0	1	0	44	0	0	0	0	0	2	0	0	0	0	0
VP16 L2	6	0	0	23	26	0	0	0	0	14	13	0	0	3	38	6
VP16 L1	1	0	0	0	0	0	0	0	0	0	57	12	0	32	18	0



VP15 L2	0	0	104	0	0	0	0	0	0	56	3	0	0	49	0	0
VP15 L1	2	0	0	0	0	0	0	0	0	61	22	7	0	22	0	17
VP14 L2	4	0	0	0	0	0	0	0	0	26	2	43	0	20	3	42
VP14 L1	17	0	21	0	0	0	0	0	0	87	0	59	0	9	0	11
VP12 L2	4	0	0	24	2	0	49	0	0	40	3	0	0	65	3	9
Vp12 L2	4	0	0	24	2	0	49	0	0	40	3	0	0	65	3	9

## Deskriptive Statistik browse

	N	Minimum	Maximum B	Mittelwert	Standardabweichung
BE-BI	30	0,0	129,0	39,367	38,7525
BE-IB	30	0,0	11,0	,667	2,2944
BE-IG	30	0,0	302,0	38,900	67,2363
LE-AA	30	0,0	585,0	221,667	174,6048
LE-EL	30	0,0	123,0	46,367	40,2334
LE-IB	30	0,0	52,0	5,433	12,9180
LE-LI	30	0,0	140,0	25,567	36,9395
LE-LK	30	0,0	40,0	4,400	10,7241
LE-UES	30	0,0	27,0	4,300	6,4549
QL-AA	30	0,0	221,0	66,567	70,4741
QL-EL	30	0,0	47,0	11,333	14,6625
QL-IB	30	0,0	38,0	5,833	10,4752
QL-IG	30	0,0	19,0	1,267	4,8205
QL-LI	30	0,0	238,0	55,733	54,2650
QL-LK	30	0,0	113,0	18,033	27,4835
QL-UES	30	0,0	226,0	46,700	48,0547
Gültige Werte (Listenweise)	30				

## Deskriptive Statistik learn

	N	Minimum	Maximum LE	Mittelwert	Standardabweichung
BE-BI	30	5	129	52,07	35,548
BE-IB	30	0	2	,07	,365
BE-IG	30	0	46	8,17	15,091
LE-AA	30	96	626	330,50	141,619
LE-EL	30	0	147	38,30	44,918
LE-IB	30	0	62	4,07	14,049
LE-LI	30	0	179	25,60	46,379
LE-LK	30	0	12	,40	2,191
LE-UES	30	0	65	5,77	12,743
QL-AA	30	0	344	77,23	71,592
QL-EL	30	0	22	4,57	7,370
QL-IB	30	0	11	,70	2,200
QL-IG	30	0	0	0,00	0,000
QL-LI	30	0	94	24,90	26,336
QL-LK	30	0	46	4,40	10,614
QL-UES	30	0	227	66,93	58,800
Gültige Werte (Listenweise)	30				

## Deskriptive Statistik look up

	N	Minimum	Maximum L	Mittelwert	Standardabweichung
BE-BI	30	0	17	2,20	3,438
BE-IB	30	0	7	,23	1,278
BE-IG	30	0	104	8,47	22,054
LE-AA	30	0	72	14,37	21,350
LE-EL	30	0	44	3,63	10,420
LE-IB	30	0	106	9,40	23,943
LE-LI	30	0	100	13,90	22,449
LE-LK	30	0	63	2,77	11,947
LE-UES	30	0	6	,57	1,278
QL-AA	30	0	134	40,70	37,566
QL-EL	30	0	125	11,90	24,522
QL-IB	30	0	69	10,37	17,468
QL-IG	30	0	0	0,00	0,000
QL-LI	30	0	144	39,53	33,382
QL-LK	30	0	53	6,60	13,206
QL-UES	30	0	105	18,67	29,953
Gültige Werte (Listenweise)	30				

## 5.4. Test auf Normalverteilung der Videolabelkombinationen

Tabelle 20-4 Test auf Normalverteilung der Häufigkeiten über alle Tasks

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
BE-BI	,198	90	,000	,810	90	,000
BE-IB	,528	90	,000	,220	90	,000
BE-IG	,336	90	,000	,473	90	,000
LE-AA	,153	90	,000	,881	90	,000
LE-EL	,229	90	,000	,761	90	,000
LE-IB	,440	90	,000	,416	90	,000
LE-LI	,277	90	,000	,644	90	,000
LE-LK	,484	90	,000	,302	90	,000
LE-UES	,338	90	,000	,459	90	,000
QL-AA	,168	90	,000	,826	90	,000
QL-EL	,298	90	,000	,560	90	,000
QL-IB	,324	90	,000	,524	90	,000
QL-IG	,537	90	,000	,132	90	,000
QL-LI	,166	90	,000	,823	90	,000
QL-LK	,309	90	,000	,576	90	,000
QL-UES	,192	90	,000	,812	90	,000

a. Signifikanzkorrektur nach Lilliefors

Tabelle 20-5 Test auf Normaverteilung Häufigkeiten casual leisure

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
BE-BI	,214	30	,001	,861	30	,001
BE-IB	,514	30	,000	,334	30	,000
BE-IG	,281	30	,000	,625	30	,000
LE-AA	,165	30	,037	,921	30	,028
LE-EL	,133	30	,183	,906	30	,012
LE-IB	,430	30	,000	,491	30	,000
LE-LI	,251	30	,000	,731	30	,000
LE-LK	,393	30	,000	,475	30	,000
LE-UES	,295	30	,000	,719	30	,000
QL-AA	,187	30	,009	,816	30	,000
QL-EL	,267	30	,000	,769	30	,000
QL-IB	,289	30	,000	,621	30	,000
QL-IG	,537	30	,000	,275	30	,000
QL-LI	,190	30	,007	,826	30	,000
QL-LK	,256	30	,000	,715	30	,000
QL-UES	,166	30	,035	,820	30	,000

a. Signifikanzkorrektur nach Lilliefors

Tabelle 20-6 Test auf Normalverteilung Häufigkeiten learn

Tests auf Normalverteilung <sup>c</sup>						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
BE-BI	,190	30	,007	,898	30	,008
BE-IB	,539	30	,000	,180	30	,000
BE-IG	,383	30	,000	,595	30	,000
LE-AA	,107	30	,200	,967	30	,451
LE-EL	,197	30	,004	,779	30	,000
LE-IB	,481	30	,000	,324	30	,000
LE-LI	,290	30	,000	,597	30	,000
LE-LK	,539	30	,000	,180	30	,000
LE-UES	,325	30	,000	,503	30	,000
QL-AA	,207	30	,002	,799	30	,000
QL-EL	,331	30	,000	,648	30	,000
QL-IB	,458	30	,000	,370	30	,000
QL-LI	,218	30	,001	,856	30	,001
QL-LK	,394	30	,000	,484	30	,000
QL-UES	,214	30	,001	,872	30	,002

\*. Dies ist eine untere Grenze der echten Signifikanz.

- a. Signifikanzkorrektur nach Lilliefors  
 c. QL-IG ist konstant und wurde weggelassen.

Tabelle 20-7 Test auf Normalverteilung Häufigkeiten look up

Tests auf Normalverteilung <sup>b</sup>						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
BE-BI	,261	30	,000	,652	30	,000
BE-IB	,539	30	,000	,180	30	,000
BE-IG	,433	30	,000	,451	30	,000
LE-AA	,349	30	,000	,717	30	,000
LE-EL	,436	30	,000	,404	30	,000
LE-IB	,423	30	,000	,462	30	,000
LE-LI	,271	30	,000	,680	30	,000
LE-LK	,525	30	,000	,251	30	,000
LE-UES	,438	30	,000	,503	30	,000
QL-AA	,139	30	,142	,903	30	,010
QL-EL	,314	30	,000	,493	30	,000
QL-IB	,276	30	,000	,645	30	,000
QL-LI	,118	30	,200	,909	30	,014
QL-LK	,341	30	,000	,576	30	,000
QL-UES	,301	30	,000	,669	30	,000

\*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

b. QL-IG ist konstant und wurde weggelassen.

## 5.5. Test auf Varianzhomogenität der drei Aufgabengruppen

Tabelle 20-8 Test auf Varianzgleichheit zwischen den Gruppen

Test der Homogenität der Varianzen				
	Levene-Statistik	df1	df2	Signifikanz
BE-BI	22,145	2	87	,000
BE-IB	5,115	2	87	,008
BE-IG	9,735	2	87	,000
LE-AA	26,466	2	87	,000
LE-EL	14,484	2	87	,000
LE-IB	2,994	2	87	,055
LE-LI	2,324	2	87	,104
LE-LK	4,687	2	87	,012
LE-UES	7,409	2	87	,001
QL-AA	3,631	2	87	,031
QL-EL	3,239	2	87	,044
QL-IB	11,541	2	87	,000
QL-IG	9,609	2	87	,000
QL-LI	4,594	2	87	,013
QL-LK	9,274	2	87	,000
QL-UES	4,112	2	87	,020

## 5.6. Deskriptive Statistik der Häufigkeiten

Tabelle 20-9 absolute Häufigkeiten casual-leisure

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
BE-BI	30	,0	129,0	39,367	38,7525
BE-IB	30	,0	11,0	,667	2,2944
BE-IG	30	,0	302,0	38,900	67,2363
LE-AA	30	,0	585,0	221,667	174,6048
LE-EL	30	,0	123,0	46,367	40,2334
LE-IB	30	,0	52,0	5,433	12,9180
LE-LI	30	,0	140,0	25,567	36,9395
LE-LK	30	,0	40,0	4,400	10,7241
LE-UES	30	,0	27,0	4,300	6,4549
QL-AA	30	,0	221,0	66,567	70,4741
QL-EL	30	,0	47,0	11,333	14,6625
QL-IB	30	,0	38,0	5,833	10,4752
QL-IG	30	,0	19,0	1,267	4,8205
QL-LI	30	,0	238,0	55,733	54,2650
QL-LK	30	,0	113,0	18,033	27,4835
QL-UES	30	,0	226,0	46,700	48,0547
Gültige Werte (Listenweise)	30				

Tabelle 20-10 absolute Häufigkeiten learn

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
BE-BI	30	5	129	52,07	35,548
BE-IB	30	0	2	,07	,365
BE-IG	30	0	46	8,17	15,091
LE-AA	30	96	626	330,50	141,619
LE-EL	30	0	147	38,30	44,918
LE-IB	30	0	62	4,07	14,049
LE-LI	30	0	179	25,60	46,379
LE-LK	30	0	12	,40	2,191
LE-UES	30	0	65	5,77	12,743
QL-AA	30	0	344	77,23	71,592
QL-EL	30	0	22	4,57	7,370
QL-IB	30	0	11	,70	2,200
QL-IG	30	0	0	,00	,000
QL-LI	30	0	94	24,90	26,336
QL-LK	30	0	46	4,40	10,614
QL-UES	30	0	227	66,93	58,800
Gültige Werte (Listenweise)	30				

Tabelle 20-11 absolute Häufigkeiten look up

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
BE-BI	30	0	17	2,20	3,438
BE-IB	30	0	7	,23	1,278
BE-IG	30	0	104	8,47	22,054
LE-AA	30	0	72	14,37	21,350
LE-EL	30	0	44	3,63	10,420
LE-IB	30	0	106	9,40	23,943
LE-LI	30	0	100	13,90	22,449
LE-LK	30	0	63	2,77	11,947
LE-UES	30	0	6	,57	1,278
QL-AA	30	0	134	40,70	37,566
QL-EL	30	0	125	11,90	24,522
QL-IB	30	0	69	10,37	17,468
QL-IG	30	0	0	,00	,000
QL-LI	30	0	144	39,53	33,382
QL-LK	30	0	53	6,60	13,206
QL-UES	30	0	105	18,67	29,953
Gültige Werte (Listenweise)	30				

Tabelle 20-12 relative Häufigkeiten casual-leisure

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
BE-BI	17	,0000000000000000	,1674491392801252	,056166899880798	,055079705757487
BE-IB	17	,0000000000000000	,0082212257100149	,000881954984668	,002502814669950
BE-IG	17	,0000000000000000	,3446105428796223	,057814508816909	,092238619744001
LE-AA	17	,0000000000000000	,7428977272727273	,393163161416434	,235725395196210
LE-EL	17	,0000000000000000	,2061855670103093	,089220051901763	,061975714691273
LE-IB	17	,0000000000000000	,1340206185567010	,012723394273899	,033481157314316
LE-LI	17	,0000000000000000	,1855670103092784	,036498279182120	,051800964198749
LE-LK	17	,0000000000000000	,0497592295345104	,006428074616276	,014962799474844
LE-UES	17	,0000000000000000	,0386597938144330	,009243700033664	,011907467088649
QL-AA	17	,0000000000000000	,3830155979202773	,121544789315067	,104873604480576
QL-EL	17	,0000000000000000	,0611620795107034	,018291101366622	,021934880130546
QL-IB	17	,0000000000000000	,0658578856152513	,009127523479000	,015882713263078
QL-IG	17	,0000000000000000	,0329289428076257	,001936996635743	,007986441725633
QL-LI	17	,0142045454545455	,2545155993431856	,085934653462564	,070321956291229
QL-LK	17	,0000000000000000	,1203407880724175	,027543745489574	,039094862325149
QL-UES	17	,0028409090909091	,2603461249059443	,073481165086075	,060781003301008
Gültige Werte (Listenweise)	17				

Tabelle 20-13 relative Häufigkeiten learn

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
BE-BI	17	,0023603461841070	,1963470319634703	,071843502588221	,050752818518232
BE-IB	17	,0000000000000000	,0015467904098995	,000090987671171	,000375151778865
BE-IG	17	,0000000000000000	,0663780663780664	,011970104755253	,019795184235821
LE-AA	17	,1461187214611872	,7641921397379912	,525195475405468	,166066496169502
LE-EL	17	,0000000000000000	,2009132420091324	,055224459425905	,053753216319630
LE-IB	17	,0000000000000000	,0461997019374069	,005586399173109	,014270479001264
LE-LI	17	,0000000000000000	,2978369384359401	,036530256347763	,071760756636109
LE-LK	17	,0000000000000000	,0088954781319496	,000523263419526	,002157470348729
LE-UES	17	,0000000000000000	,0611065235342692	,008088092856238	,015725158815961
QL-AA	17	,0317460317460317	,4146079484425349	,127951394130304	,093640057668131
QL-EL	17	,0000000000000000	,0363036303630363	,006582580340961	,009684909197981
QL-IB	17	,0000000000000000	,0338316286388670	,002903629419482	,008254454950039
QL-IG	17	,0000000000000000	,0267505900865460	,001573564122738	,006487971086731
QL-LI	17	,0000000000000000	,0797317436661699	,035742694873930	,029222175406821
QL-LK	17	,0000000000000000	,0440835266821346	,005934545784285	,011999010041391
QL-UES	17	,0000000000000000	,2838522809558291	,104259049685647	,075463222094748
Gültige Werte (Listenweise)	17				

Tabelle 20-14 relative Häufigkeiten look up

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
BE-BI	17	,0000000000000000	,0610465116279070	,010125332278433	,015234927146090
BE-IB	17	,0000000000000000	,0108191653786708	,000636421492863	,002624033037487
BE-IG	17	,0000000000000000	,4719101123595506	,055584941722123	,131051385846920
LE-AA	17	,0000000000000000	,3016627078384798	,072069104954236	,090220727406722
LE-EL	17	,0000000000000000	,2587412587412587	,034187862348724	,071598457456216
LE-IB	17	,0000000000000000	,2379912663755458	,040786447898954	,080777568940899
LE-LI	17	,0000000000000000	,2539682539682540	,077487213536846	,091748373899258
LE-LK	17	,0000000000000000	,1496437054631829	,011714626821072	,037509187642663
LE-UES	17	,0000000000000000	,0198019801980198	,003059928660983	,006133207293588
QL-AA	17	,0000000000000000	,7872340425531915	,243385852948990	,194553540581240
QL-EL	17	,0000000000000000	,2811244979919679	,073784387730685	,079637546082281
QL-IB	17	,0000000000000000	,2965116279069768	,052100119268297	,070843137010275
QL-IG	17	0	0	,00	,000
QL-LI	17	,0000000000000000	,3272727272727273	,212497656298841	,093593966131822
QL-LK	17	,0000000000000000	,2248995983935743	,036592056674213	,064347017594858
QL-UES	17	,0000000000000000	,3079710144927536	,075988047364741	,086203709807268
Gültige Werte (Listenweise)	17				

Tabelle 20-15 Mittelwerte der relativen Häufigkeiten für die Videolabelkombinationen, alle TT

	BE-BI	BE-IB	BE-IG	LE-AA	LE-EL	LE-IB	LE-LI	LE-LK	LE-UES	QL-AA	QL-EL	QL-IB	QL-IG	QL-LI	QL-LK	QL-UES
Browse	0,0562	0,0009	0,0578	0,3932	0,0892	0,0127	0,0365	0,0064	0,0092	0,1215	0,0183	0,0091	0,0019	0,0859	0,0275	0,0735
Learn	0,0718	0,0001	0,0120	0,5252	0,0552	0,0056	0,0365	0,0005	0,0081	0,1280	0,0066	0,0029	0,0016	0,0357	0,0059	0,1043
Look up	0,0101	0,0006	0,0556	0,0721	0,0342	0,0408	0,0775	0,0117	0,0031	0,2434	0,0738	0,0521	0,0001	0,2125	0,0366	0,0760

## 5.7. Einfluss der emotionalen Lage auf die Verwendung der Inhaltselemente

Tabelle 20-16 T-test über Kategorie arousal des SAM, Trennwert 4,  $n < 4 = 41$ ,  $n > 4 = 48$ 

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
BE-BI	Varianzen sind gleich	,270	,605	-,006	87	,996	-,044	7,893	-15,732	15,644
	Varianzen sind nicht gleich			-,005	81,435	,996	-,044	7,958	-15,876	15,789
BE-IB	Varianzen sind gleich	2,923	,091	,875	87	,384	,288	,329	-,366	,941
	Varianzen sind nicht gleich			,909	78,087	,366	,288	,316	-,342	,917
BE-IG	Varianzen sind gleich	8,978	,004	-	87	,154	-13,364	9,295	-31,840	5,111
	Varianzen sind nicht gleich			1,349	47,464	,184	-13,364	9,910	-33,295	6,567
LE-AA	Varianzen sind gleich	,092	,762	,236	87	,814	9,299	39,414	-69,041	87,639
	Varianzen sind nicht gleich			,236	85,501	,814	9,299	39,329	-68,891	87,489
LE-EL	Varianzen sind gleich	,144	,705	-,494	87	,622	-4,191	8,478	-21,042	12,659
	Varianzen sind nicht gleich			-,494	84,571	,623	-4,191	8,484	-21,062	12,680

LE-IB	Varianzen sind gleich	6,441	,013	-	87	,177	-5,101	3,749	-12,552	2,350
	Varianzen sind nicht gleich			-	79,078	,183	-5,101	3,796	-12,658	2,455
LE-LI	Varianzen sind gleich	,189	,665	-	87	,758	-2,431	7,855	-18,043	13,181
	Varianzen sind nicht gleich			-	86,175	,756	-2,431	7,816	-17,969	13,106
LE-LK	Varianzen sind gleich	8,016	,006	-	87	,130	-3,049	1,993	-7,011	,913
	Varianzen sind nicht gleich			-	56,173	,151	-3,049	2,093	-7,241	1,143
LE-UES	Varianzen sind gleich	1,156	,285	-	87	,263	-2,037	1,808	-5,632	1,557
	Varianzen sind nicht gleich			-	61,367	,284	-2,037	1,883	-5,802	1,728
QL-AA	Varianzen sind gleich	3,058	,084	-	87	,184	17,940	13,395	-8,684	44,564
	Varianzen sind nicht gleich			-	82,589	,171	17,940	12,996	-7,910	43,790
QL-EL	Varianzen sind gleich	1,616	,207	-	87	,882	-5,551	3,690	-7,886	6,784
	Varianzen sind nicht gleich			-	59,048	,887	-5,551	3,857	-8,268	7,166
QL-IB	Varianzen sind gleich	,503	,480	-	87	,739	,885	2,652	-4,386	6,155
	Varianzen sind nicht gleich			-	86,569	,738	,885	2,633	-4,349	6,118
QL-IG	Varianzen sind gleich	7,618	,007	-	87	,190	,792	,600	-,400	1,984
	Varianzen sind nicht gleich			-	47,000	,159	,792	,554	-,322	1,906
QL-LI	Varianzen sind gleich	,391	,533	-	87	,409	7,339	8,852	-10,254	24,933
	Varianzen sind nicht gleich			-	86,485	,401	7,339	8,689	-9,933	24,611
QL-LK	Varianzen sind gleich	3,433	,067	-	87	,308	-4,240	4,139	-12,466	3,986
	Varianzen sind nicht gleich			-	62,987	,328	-4,240	4,299	-12,830	4,350
QL-UES	Varianzen sind gleich	2,596	,111	-	87	,033	22,857	10,571	1,845	43,868
	Varianzen sind nicht gleich			-	83,669	,029	22,857	10,281	2,411	43,302



Tabelle 20-17 T-test über Kategorie valences des SAM, Trennwert 3,  $n \leq 3 = 64$ ,  $n > 3 = 25$ 

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
BE-BI	Varianzen sind gleich	,000	,997	,222	87	,825	1,944	8,751	-15,450	19,338
	Varianzen sind nicht gleich			,213	40,494	,832	1,944	9,116	-16,473	20,361
BE-IB	Varianzen sind gleich	6,773	,011	1,249	87	,215	-,453	,363	-1,174	,268
	Varianzen sind nicht gleich			2,005	63,000	,049	-,453	,226	-,905	-,002
BE-IG	Varianzen sind gleich	6,509	,012	1,613	87	,110	-16,575	10,278	-37,004	3,854
	Varianzen sind nicht gleich			2,409	81,364	,018	-16,575	6,882	-30,267	-2,883
LE-AA	Varianzen sind gleich	,974	,326	-,823	87	,413	-35,831	43,558	122,407	50,744
	Varianzen sind nicht gleich			-,859	48,109	,395	-35,831	41,705	119,680	48,017
LE-EL	Varianzen sind gleich	3,256	,075	1,575	87	,119	14,624	9,284	-3,829	33,077
	Varianzen sind nicht gleich			1,410	35,871	,167	14,624	10,369	-6,408	35,655
LE-IB	Varianzen sind gleich	2,689	,105	,767	87	,445	3,211	4,187	-5,111	11,534
	Varianzen sind nicht gleich			,616	30,746	,542	3,211	5,210	-7,418	13,840
LE-LI	Varianzen sind gleich	,627	,431	-,430	87	,668	-3,744	8,707	-21,050	13,561
	Varianzen sind nicht gleich			-,473	54,335	,638	-3,744	7,908	-19,597	12,108
LE-LK	Varianzen sind gleich	,234	,630	-,243	87	,809	-,543	2,240	-4,995	3,908
	Varianzen sind nicht gleich			-,266	54,066	,791	-,543	2,038	-4,630	3,543
LE-UES	Varianzen sind gleich	,080	,778	,121	87	,904	,244	2,020	-3,771	4,260
	Varianzen sind nicht gleich			,145	67,070	,885	,244	1,685	-3,118	3,607
QL-AA	Varianzen sind gleich	,014	,905	1,118	87	,267	16,654	14,901	-12,964	46,273
	Varianzen sind nicht gleich			1,213	52,491	,231	16,654	13,733	-10,897	44,206
QL-EL	Varianzen sind gleich	2,074	,153	-,756	87	,452	-3,083	4,080	-11,193	5,027
	Varianzen sind nicht gleich			1,008	83,716	,317	-3,083	3,060	-9,168	3,001
QL-IB	Varianzen sind gleich	7,938	,006	1,423	87	,158	-4,139	2,909	-9,921	1,642
	Varianzen sind nicht gleich			2,080	84,657	,041	-4,139	1,990	-8,097	-,182
QL-IG	Varianzen sind gleich	3,367	,070	-,888	87	,377	-,594	,669	-1,923	,735
	Varianzen sind nicht gleich			1,426	63,000	,159	-,594	,416	-1,426	,239
QL-LI	Varianzen sind gleich	,167	,683	,576	87	,566	5,671	9,837	-13,882	25,223
	Varianzen sind nicht gleich			,589	45,887	,559	5,671	9,626	-13,707	25,048
QL-LK	Varianzen sind gleich	6,660	,012	1,782	87	,078	8,084	4,536	-,930	17,099
	Varianzen sind nicht gleich			1,390	29,644	,175	8,084	5,815	-3,797	19,966
QL-UES	Varianzen sind gleich	,350	,556	1,422	87	,159	16,916	11,897	-6,731	40,564
	Varianzen sind nicht gleich			1,396	42,268	,170	16,916	12,120	-7,538	41,370

### 5.8. Kruskal- Wallis Test über die ersten 50 frames

Tabelle 20-18 Test über die Gleichverteilung der Videolabelkombinationen in den ersten 50 x 50 ms, absolute Häufigkeiten

Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von BE-BI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,108	Nullhypothese behalten.
2	Die Verteilung von BE-IB ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,368	Nullhypothese behalten.
3	Die Verteilung von BE-IG ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,396	Nullhypothese behalten.
4	Die Verteilung von LE-AA ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,015	Nullhypothese ablehnen.
5	Die Verteilung von LE-EL ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,057	Nullhypothese behalten.
6	Die Verteilung von LE-IB ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,317	Nullhypothese behalten.
7	Die Verteilung von LE-LI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,699	Nullhypothese behalten.
8	Die Verteilung von LE-LK ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,599	Nullhypothese behalten.
9	Die Verteilung von LE-UES ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,378	Nullhypothese behalten.
10	Die Verteilung von QL-AA ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,927	Nullhypothese behalten.
11	Die Verteilung von QL-EL ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,027	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
12	Die Verteilung von QL-IB ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,183	Nullhypothese behalten.
13	Die Verteilung von QL-IG ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	1,000	Nullhypothese behalten.
14	Die Verteilung von QL-LI ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,570	Nullhypothese behalten.
15	Die Verteilung von QL-LK ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,759	Nullhypothese behalten.
16	Die Verteilung von QL-UES ist über Kategorien von V1 gleich.	Kruskal-Wallis-Test unabhängiger Stichproben	,018	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,05.

## 5.9. Einfluss des Geschlechts auf das Suchverhalten

Tabelle 20-19 t-Test Männlich vs. weiblich VP, relative Häufigkeiten

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
BE-BI	Varianzen sind gleich	2,486	,118	,809	88	,420	6,378	7,879	-9,280	22,035
	Varianzen sind nicht gleich			,840	87,455	,403	6,378	7,594	-8,716	21,471
BE-IB	Varianzen sind gleich	,965	,329	-,520	88	,604	-,171	,329	-,825	,483
	Varianzen sind nicht gleich			-,521	80,162	,604	-,171	,329	-,825	,483
BE-IG	Varianzen sind gleich	1,771	,187	1,581	88	,117	-14,647	9,264	-33,057	3,763
	Varianzen sind nicht gleich			1,477	58,078	,145	-14,647	9,915	-34,493	5,200
LE-AA	Varianzen sind gleich	,392	,533	,687	88	,494	27,104	39,473	-51,340	105,548
	Varianzen sind nicht gleich			,696	83,459	,488	27,104	38,947	-50,353	104,561
LE-EL	Varianzen sind gleich	1,041	,310	1,708	88	,091	14,277	8,361	-2,338	30,893
	Varianzen sind nicht gleich			1,739	84,588	,086	14,277	8,208	-2,044	30,598
LE-IB	Varianzen sind gleich	3,096	,082	1,135	88	,259	-4,263	3,756	-11,726	3,200
	Varianzen sind nicht gleich			1,120	75,787	,266	-4,263	3,806	-11,844	3,317
LE-LI	Varianzen sind gleich	6,856	,010	1,420	88	,159	11,030	7,769	-4,409	26,470
	Varianzen sind nicht gleich			1,562	78,866	,122	11,030	7,064	-3,030	25,090
LE-LK	Varianzen sind gleich	,000	,990	,042	88	,967	,084	2,016	-3,923	4,091
	Varianzen sind nicht gleich			,043	87,624	,966	,084	1,940	-3,771	3,939
LE-UES	Varianzen sind gleich	14,843	,000	2,578	88	,012	-4,523	1,754	-8,010	-1,037
	Varianzen sind nicht gleich			2,245	40,699	,030	-4,523	2,015	-8,593	-,453
QL-AA	Varianzen sind gleich	,552	,460	-,997	88	,321	-13,436	13,475	-40,214	13,342
	Varianzen sind nicht gleich			-,963	68,753	,339	-13,436	13,950	-41,267	14,395
QL-EL	Varianzen sind gleich	,371	,544	-,394	88	,694	-1,451	3,683	-8,770	5,867
	Varianzen sind nicht gleich			-,368	57,725	,714	-1,451	3,946	-9,351	6,448
QL-IB	Varianzen sind gleich	24,143	,000	2,149	88	,034	-5,554	2,584	-10,688	-,419
	Varianzen sind nicht gleich			1,879	41,501	,067	-5,554	2,955	-11,519	,412
QL-IG	Varianzen sind gleich	12,667	,001	1,681	88	,096	-1,000	,595	-2,182	,182
	Varianzen sind nicht gleich			1,434	37,000	,160	-1,000	,697	-2,413	,413
QL-LK	Varianzen sind gleich	11,269	,001	2,453	88	,016	9,872	4,025	1,873	17,872
	Varianzen sind nicht gleich			2,687	80,339	,009	9,872	3,675	2,560	17,185
QL-UES	Varianzen sind gleich	4,227	,043	1,312	88	,193	-14,129	10,770	-35,532	7,275
	Varianzen sind nicht gleich			1,241	62,021	,219	-14,129	11,382	-36,880	8,623

## 5.10. Einfluss des Vorwissens auf die Verwendung der Inhaltselemente

Tabelle 20-20 t-Test Gruppe mit keinem oder wenig Vorwissen vs. Vorwissen

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
BE-BI	Varianzen sind gleich	1,523	,222	-	57	,159	-13,496	9,461	-32,442	5,450
BE-BI	Varianzen sind nicht gleich			-	53,700	,141	-13,496	9,031	-31,605	4,612
BE-IB	Varianzen sind gleich	10,824	,002	1,426	57	,121	,391	,249	-,107	,889
BE-IB	Varianzen sind nicht gleich			1,494	22,000	,224	,391	,313	-,257	1,040
BE-IG	Varianzen sind gleich	3,749	,058	1,573	57	,344	4,808	5,040	-5,284	14,900
BE-IG	Varianzen sind nicht gleich			,954	31,219	,402	4,808	5,654	-6,721	16,337
LE-AA	Varianzen sind gleich	3,755	,058	,850	57	,140	-72,543	48,523	-	24,623
LE-AA	Varianzen sind nicht gleich			1,495	54,608	,120	-72,543	45,947	169,709	-
LE-EL	Varianzen sind gleich	,007	,932	1,579	57	,764	-2,995	9,945	-22,909	16,919
LE-EL	Varianzen sind nicht gleich			-,301	50,048	,760	-2,995	9,750	-22,578	16,588
LE-IB	Varianzen sind gleich	,307	,582	-,307	57	,816	-1,246	5,328	-11,915	9,423
LE-IB	Varianzen sind nicht gleich			-,234	54,758	,806	-1,246	5,038	-11,344	8,851
LE-LI	Varianzen sind gleich	5,769	,020	-,247	57	,055	-18,808	9,598	-38,027	,411
LE-LI	Varianzen sind nicht gleich			1,960	47,371	,024	-18,808	8,059	-35,018	-2,598
LE-LK	Varianzen sind gleich	,157	,694	2,334	57	,878	-,359	2,334	-5,033	4,315
LE-LK	Varianzen sind nicht gleich			-,154	52,518	,859	-,359	2,012	-4,396	3,678
LE-UES	Varianzen sind gleich	,507	,479	-,178	57	,673	-1,074	2,534	-6,148	4,000
LE-UES	Varianzen sind nicht gleich			-,424	55,491	,633	-1,074	2,234	-5,549	3,402
QL-AA	Varianzen sind gleich	1,026	,315	-,481	57	,049	31,378	15,588	,164	62,593
QL-AA	Varianzen sind nicht gleich			1,845	34,435	,074	31,378	17,009	-3,173	65,929
QL-EL	Varianzen sind gleich	6,590	,013	1,821	57	,074	8,795	4,831	-,878	18,468
QL-EL	Varianzen sind nicht gleich			1,502	24,360	,146	8,795	5,857	-3,285	20,874
QL-IB	Varianzen sind gleich	,120	,730	,249	57	,804	,896	3,597	-6,306	8,099
QL-IB	Varianzen sind nicht gleich			,240	41,302	,811	,896	3,733	-6,641	8,433
QL-LI	Varianzen sind gleich	,124	,726	2,012	57	,049	16,181	8,042	,078	32,285
QL-LI	Varianzen sind nicht gleich			2,092	52,737	,041	16,181	7,735	,665	31,698
QL-LK	Varianzen sind gleich	8,123	,006	-	57	,274	-3,537	3,200	-9,945	2,870
QL-LK	Varianzen sind nicht gleich			1,105	47,217	,194	-3,537	2,685	-8,938	1,864
QL-UES	Varianzen sind gleich	,988	,324	1,317	57	,399	-11,973	14,102	-40,213	16,266
QL-UES	Varianzen sind nicht gleich			-,849	54,416	,375	-11,973	13,377	-38,789	14,842



LE-UE S	Signifikanz (2-seitig)	,638	,594	,714	,457	,329	,480	,206		,303	,106	,371	,414	,703	,180	,026	,506
	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
QL-AA	Korrelation nach Pearson	,084	,084	,070	,038	,458	,248	,153	,110	1	,174	,015	,049	,187	,001	,046	,031
	Signifikanz (2-seitig)	,429	,434	,513	,721	,000	,018	,150	,303		,102	,889	,646	,078	,990	,669	,775
QL-EL	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	Korrelation nach Pearson	,206	,276	,104	,189	,132	,085	,171	,172	,174	1	,175	,144	,383	,216	,039	,194
QL-IB	Signifikanz (2-seitig)	,051	,008	,330	,074	,215	,424	,106	,106	,102		,100	,175	,000	,041	,714	,067
	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
QL-IG	Korrelation nach Pearson	,173	,004	,007	,241	,098	,081	,066	,095	,015	,175	1	,502	,210	,348	,020	,095
	Signifikanz (2-seitig)	,103	,967	,949	,022	,358	,449	,540	,371	,889	,100		,000	,047	,001	,848	,372
QL-LI	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	Korrelation nach Pearson	,193	,072	,059	,272	,201	,021	,166	,087	,049	,144	,502	1	,397	,171	,073	,026
QL-LK	Signifikanz (2-seitig)	,068	,500	,578	,009	,058	,847	,118	,414	,646	,175	,000		,000	,107	,493	,807
	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
QL-UE S	Korrelation nach Pearson	,054	,032	,033	,073	,113	,011	,069	,041	,187	,383	,210	,397	1	,168	,076	,045
	Signifikanz (2-seitig)	,611	,766	,758	,494	,290	,917	,518	,703	,078	,000	,047	,000		,112	,478	,672
QL-UE S	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	Korrelation nach Pearson	,015	,127	,002	,330	,038	,053	,048	,143	,001	,216	,348	,171	,168	1	,219	,250
QL-UE S	Signifikanz (2-seitig)	,885	,235	,986	,001	,722	,623	,650	,180	,990	,041	,001	,107	,112		,038	,017
	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
QL-UE S	Korrelation nach Pearson	,084	,018	,027	,158	,121	,004	,125	,235	,046	,039	,020	,073	,076		,219	,042
	Signifikanz (2-seitig)	,432	,864	,799	,136	,255	,971	,242	,026	,669	,714	,848	,493	,478	,038		,693
QL-UE S	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	Korrelation nach Pearson	,380	,179	,062	,112	,038	,113	,111	,071	,031	,194	,095	,026	,045	,250	,042	1
QL-UE S	Signifikanz (2-seitig)	,000	,091	,560	,292	,725	,287	,299	,506	,775	,067	,372	,807	,672	,017	,693	
	N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

## 6. Aufgaben aus Studie zwei

*Tabelle 20-22 Aufgaben Studie WS 2012/2013*

### **Look up:**

**LU1)** Am Wochenende wollen Sie mit ihrer Familie einen Ausflug auf die Zugspitze machen. Sie haben sich dazu entschieden, mit der Deutschen Bahn bis nach Garmisch-Partenkirchen zu fahren und von dort die historische Zahnradbahn zum Gipfel hinauf zunehmen. Um den zeitlichen Ablauf des Tages zu planen, wollen Sie in der Wikipedia nachschauen, wie lange eine Fahrt mit der Zahnradbahn dauert.

**LU3)** Sie schreiben eine Seminararbeit zum Dreißigjährigen-Krieg. Da Sie in Regensburg leben, möchten Sie wissen, ob dieser auch Spuren in ihrer Heimatstadt hinterlassen hat. Sie schauen deshalb in der Wikipedia nach, wann Regensburg im Dreißigjährigen Krieg erobert wurde

---

### **Learn:**

**LE1)** In einigen Wochen wollen Sie mit ihren Freunden durch Südostasien reisen. Auf dem Tourplan ist auch ein Aufenthalt in Kambodscha vorgesehen. In einer Fernsehreportage haben Sie einige beeindruckende Bilder der berühmten Tempelanlagen von Angkor gesehen und wollen diese unbedingt besichtigen. Um auch ihre Freunde davon zu überzeugen, dass diese Anlage einen Besuch wert ist, versuchen Sie in der Wikipedia möglichst viele interessante Informationen darüber zu finden.

**LE3)** Immer wenn ihre befreundete japanische Austauschstudentin auf die Toilette geht, dreht diese die Musik laut auf, sodass sich eine Nachbarin über die hohe Lautstärke beschwert. Da es bereits mehrere Missverständnisse aufgrund kultureller Differenzen zwischen den beiden Damen gegeben hat, möchten Sie als Vermittler fungieren. Hierzu rufen Sie die Wikipedia auf, um sich über die Toiletten- und Sanitärkultur in Japan zu informieren und ihr Verhalten besser verstehen zu können.

---

### **Casual leisure:**

**CL1)** Bitte benutzen Sie die Funktion "zufälliger Artikel" der Wikipedia. Ausgehend von dem angezeigten Artikel können Sie nun über die Hyperlinks frei zu Themen navigieren, die für Sie PERSÖNLICH interessant sind.

**CL2)** Bitte benutzen Sie die Funktion "zufälliger Artikel" der Wikipedia. Ausgehend

von dem angezeigten Artikel können Sie nun über die Hyperlinks frei zu Themen navigieren, die für Sie PERSÖNLICH interessant sind.

## 7. Fragebögen aus Studie zwei

### 7.1. Pre Study Fragebogen

---

7. Jahrgang:

8. Geschlecht: weiblich ☐ männlich ☐

9. Studium: ja ☐ nein ☐

Falls ja: Studienfach/-fächer: \_\_\_\_\_

10. Höchster Bildungsabschluss :

☐ Qualifizierter Hauptschulabschluss

☐ Realschulabschluss

☐ allgemeine Hochschulreife

☐ Fachabitur

☐ BA

☐ MA

☐ Dr.

☐ andere: \_\_\_\_\_

---

11. Wie schätzen Sie ihre Kenntnisse ein im Umgang mit der Wikipedia ein?

unerfahren

0 1 2 3 4 5

sehr erfahren

---

12. Wie stark treffen folgende Behauptungen auf Sie zu ( 0 = trifft gar nicht zu, 5 = trifft voll zu)?

6a) Ich halte die Wikipedia für eine seriöse Quelle

0 1 2 3 4 5

6b) Ich nutze die Wikipedia zum Zeitvertreib

0 1 2 3 4 5

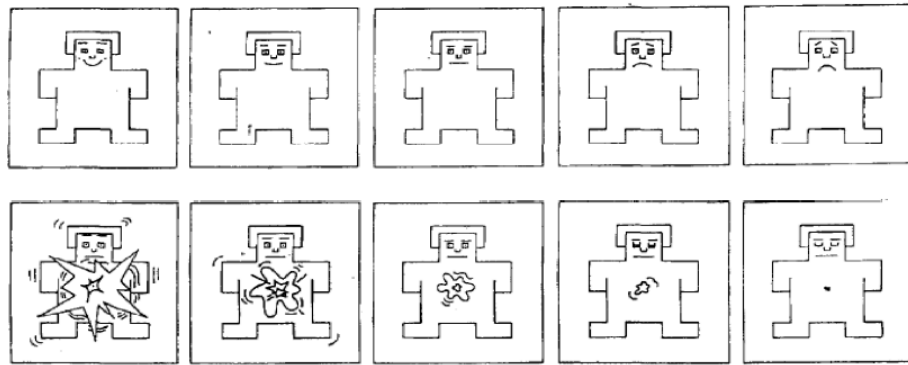
6c) Ich nutze die Wikipedia für die Arbeit

0 1 2 3 4 5



Measuring Emotion

51



7.

Beschreiben Sie Ihren momentanen Gefühlszustand durch Auswahl jeweils eines der Männchen pro Reihe!

## 7.2. Pre Task Fragebogen

1. Wie zutreffend sind folgende Aussagen (0 = trifft gar nicht zu, 5 = trifft voll zu):

1a) Ich erwarte, bei Wikipedia die Informationen in hoher Qualität zu finden

0 1 2 3 4 5

1b) Ich erwarte, bei Wikipedia die gewünschten Information in der gewünschten Form (Text, Bild, Grafik etc.) zu finden

0 1 2 3 4 5

2. Wie gut kennen Sie sich im Themengebiet des Suchtasks aus?

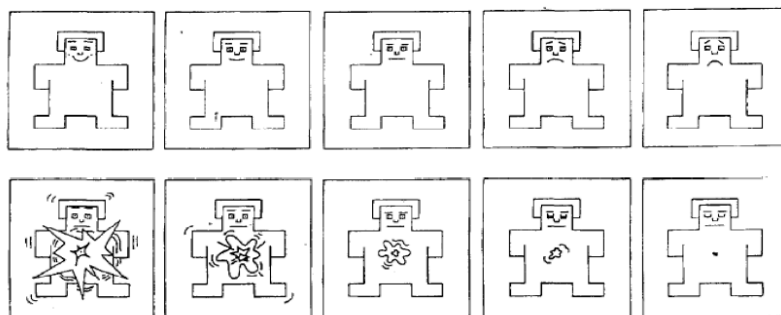
Gar nicht 0 1 2 3 4 5 Sehr gut

## 7.3. Post Task Fragebogen

1. Geben Sie ihren momentanen Gefühlszustand an:

Measuring Emotion

51



- 
1. Wie zutreffend sind folgende Aussagen (0 = trifft gar nicht zu, 5 = trifft voll zu):

1a) Ich habe die gewünschte Information  
im Text gefunden:

0	1	2	3	4	5
---	---	---	---	---	---

1b) Ich habe die gewünschte Information  
in Bildern gefunden:

0	1	2	3	4	5
---	---	---	---	---	---

1c) Ich habe die gewünschte Information  
in Tabellen oder Grafiken gefunden:

0	1	2	3	4	5
---	---	---	---	---	---

- 
2. Die Präsentation der Informationen war für mein Informationsbedürfnis angemessen

Nicht angemessen 0 1 2 3 4 5 angemessen

- 
3. Wie zutreffend sind folgende Aussagen (0 = trifft gar nicht zu, 5 = trifft voll zu):

3a) Ich hätte mir mehr Text gewünscht

0	1	2	3	4	5
---	---	---	---	---	---

3b) Ich hätte mir mehr Bilder gewünscht

0	1	2	3	4	5
---	---	---	---	---	---

3c) Ich hätte mir mehr Grafiken und Ta-  
bellen gewünscht

0	1	2	3	4	5
---	---	---	---	---	---

- 
4. Ich habe die gewünschten Informationen (hinsichtlich des Informationsbedürfnisses)  
gefunden  
(nicht für Task #...)

Nicht gefunden 0 1 2 3 4 5 Alles gefunden

- 
5. Ich bin mit dem Suchverlauf zufrieden

Nicht zufrieden 0 1 2 3 4 5 Sehr zufrieden

---

---

## 8. Daten aus Studie zwei

Für die statistische Analyse wurden so weit nicht anders angegeben jeweils Microsoft Office Excel 2007 und 54 IBM SPSS Statistics 21 verwendet

### 8.1. Lageparameter und Test auf Normalverteilung den Angaben FB

Tabelle 20-23 deskriptive Statistik post Task Fragebogen

Deskriptive Statistik			Statistik	Standardfehler
gewünschte Information in Foto	Mittelwert		3,554	,1163
	95% Konfidenzintervall des Mittelwerts	Untergrenze	3,324	
		Obergrenze	3,783	
	5% getrimmtes Mittel		3,615	
	Median		4,000	
	Varianz		2,273	
	Standardabweichung		1,5075	
	Minimum		1,0	
	Maximum		5,0	
	Spannweite		4,0	
	Interquartilbereich		3,0	
	Schiefe		-,449	,187
	Kurtosis		-1,334	,373
	Mittelwert		2,470	,1133
	95% Konfidenzintervall des Mittelwerts	Untergrenze	2,247	
gewünschte Information in Text		Obergrenze	2,694	
	5% getrimmtes Mittel		2,411	
	Median		2,000	
	Varianz		2,155	
	Standardabweichung		1,4679	
	Minimum		1,0	
	Maximum		5,0	
	Spannweite		4,0	
	Interquartilbereich		3,0	
	Schiefe		,594	,187
	Kurtosis		-1,065	,373
	Mittelwert		4,256	,0853
	95% Konfidenzintervall des Mittelwerts	Untergrenze	4,088	
		Obergrenze	4,424	
	5% getrimmtes Mittel		4,380	
gewünschte Info in Graphik/Tabelle	Median		5,000	
	Varianz		1,222	
	Standardabweichung		1,1052	
	Minimum		1,0	
	Maximum		5,0	
	Spannweite		4,0	
	Interquartilbereich		1,0	
	Schiefe		-1,438	,187
	Kurtosis		1,164	,373
	Mittelwert		3,482	,0901
	95% Konfidenzintervall des Mittelwerts	Untergrenze	3,304	
		Obergrenze	3,660	
	5% getrimmtes Mittel		3,536	
	Median		4,000	
	Varianz		1,365	
mehr Text gewünscht	Standardabweichung		1,1683	
	Minimum		1,0	
	Maximum		5,0	
	Spannweite		4,0	
	Interquartilbereich		1,0	
	Schiefe		-,720	,187
	Kurtosis		-,238	,373
	Mittelwert		3,345	,0933
	95% Konfidenzintervall des Mittelwerts	Untergrenze	3,161	
		Obergrenze	3,529	

mehr Graphiken/Tabellen gewünscht	5% getrimmtes Mittel		3,384	
	Median		3,000	
	Varianz		1,461	
	Standardabweichung		1,2087	
	Minimum		1,0	
	Maximum		5,0	
	Spannweite		4,0	
	Interquartilbereich		2,0	
	Schiefe		-,157	,187
	Kurtosis		-1,013	,373
	Mittelwert		3,530	,0966
	95% Konfidenzintervall des Mittelwerts	Untergrenze	3,339	
		Obergrenze	3,720	
	5% getrimmtes Mittel		3,586	
	Median		4,000	
	Varianz		1,568	
	Standardabweichung		1,2522	
	Minimum		1,0	
	Maximum		5,0	
	Spannweite		4,0	
	Interquartilbereich		3,0	
	Schiefe		-,365	,187
	Kurtosis		-1,123	,373

Tabelle 20-24 Test auf Normalverteilung Angaben Post Task Fragebogen

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
gewünschte Information in Foto	,272	168	,000	,810	168	,000
gewünschte Information in Text	,227	168	,000	,829	168	,000
gewünschte Info in Graphik/Tabelle	,351	168	,000	,705	168	,000
mehr Text gewünscht	,266	168	,000	,866	168	,000
mehr Fotos gewünscht	,176	168	,000	,901	168	,000
mehr Graphiken/Tabellen gewünscht	,224	168	,000	,868	168	,000

a. Signifikanzkorrektur nach Lilliefors

## 8.2. Deskriptive Statistik der absoluten Häufigkeiten

Tabelle 20-25 Videolabels/Nutzeraktionen deskriptive Statistik

Deskriptive Statistik <sup>a,b,c</sup>				
	V1		Statistik	Standardfehler
RE-LI	1	Mittelwert	13,50	4,959
		95% Konfidenzintervall des Mittelwerts	3,19	
		Untergrenze	23,81	
		Obergrenze		
		5% getrimmtes Mittel	9,80	
		Median	,00	
		Varianz	540,929	
		Standardabweichung	23,258	
		Minimum	0	
		Maximum	99	
		Spannweite	99	
		Interquartilbereich	21	
		Schiefe	2,631	,491
		Kurtosis	8,375	,953

		Mittelwert		5,85	1,900
		95% Konfidenzintervall des Mittelwerts	Untergrenze	1,93	
			Obergrenze	9,76	
		5% getrimmtes Mittel		4,56	
		Median		,00	
		Varianz		93,815	
	2	Standardabweichung		9,686	
		Minimum		0	
		Maximum		36	
		Spannweite		36	
		Interquartilbereich		11	
		Schiefe		2,039	,456
		Kurtosis		3,928	,887
		Mittelwert		6,67	3,535
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,79	
			Obergrenze	14,13	
	3	5% getrimmtes Mittel		3,85	
		Median		,00	
		Varianz		224,941	
		Standardabweichung		14,998	
		Minimum		0	
		Maximum		64	
		Spannweite		64	
		Interquartilbereich		8	
		Schiefe		3,648	,536
		Kurtosis		14,326	1,038
	1	Mittelwert		7,91	1,959
		95% Konfidenzintervall des Mittelwerts	Untergrenze	3,83	
			Obergrenze	11,98	
		5% getrimmtes Mittel		7,18	
		Median		6,00	
		Varianz		84,468	
		Standardabweichung		9,191	
		Minimum		0	
		Maximum		29	
		Spannweite		29	
RE-IN	1	Interquartilbereich		15	
		Schiefe		,996	,491
		Kurtosis		,238	,953
		Mittelwert		38,88	4,302
		95% Konfidenzintervall des Mittelwerts	Untergrenze	30,02	
			Obergrenze	47,75	
		5% getrimmtes Mittel		37,84	
		Median		32,50	
		Varianz		481,226	
	2	Standardabweichung		21,937	
		Minimum		7	
		Maximum		88	
		Spannweite		81	
		Interquartilbereich		29	
		Schiefe		,931	,456
		Kurtosis		,216	,887
		Mittelwert		42,22	9,043
		95% Konfidenzintervall des Mittelwerts	Untergrenze	23,14	
			Obergrenze	61,30	
	3	5% getrimmtes Mittel		39,69	
		Median		24,00	
		Varianz		1472,065	
		Standardabweichung		38,368	
		Minimum		6	
		Maximum		124	
		Spannweite		118	
		Interquartilbereich		74	
		Schiefe		,969	,536
		Kurtosis		-,522	1,038
RE-TX	1	Mittelwert		45,91	9,698
		95% Konfidenzintervall des Mittelwerts	Untergrenze	25,74	
			Obergrenze	66,08	
		5% getrimmtes Mittel		43,25	
		Median		26,00	
		Varianz		2069,229	

			Standardabweichung		45,489	
			Minimum		0	
			Maximum		141	
			Spannweite		141	
			Interquartilbereich		88	
			Schiefe		,795	,491
			Kurtosis		-,794	,953
			Mittelwert		164,88	18,156
			95% Konfidenzintervall des Mit-	Untergrenze	127,49	
			telwerts	Obergrenze	202,28	
			5% getrimmtes Mittel		164,76	
			Median		168,00	
			Varianz		8570,186	
	2		Standardabweichung		92,575	
			Minimum		0	
			Maximum		343	
			Spannweite		343	
			Interquartilbereich		138	
			Schiefe		-,158	,456
			Kurtosis		-,597	,887
			Mittelwert		106,72	17,203
			95% Konfidenzintervall des Mit-	Untergrenze	70,43	
			telwerts	Obergrenze	143,02	
			5% getrimmtes Mittel		102,64	
			Median		84,00	
	3		Varianz		5327,036	
			Standardabweichung		72,987	
			Minimum		0	
			Maximum		287	
			Spannweite		287	
			Interquartilbereich		62	
			Schiefe		1,263	,536
			Kurtosis		1,433	1,038
			Mittelwert		3,23	1,105
			95% Konfidenzintervall des Mit-	Untergrenze	,93	
			telwerts	Obergrenze	5,52	
			5% getrimmtes Mittel		2,52	
RE-HL	1		Median		,00	
			Varianz		26,851	
			Standardabweichung		5,182	
			Minimum		0	
			Maximum		20	
			Spannweite		20	
			Interquartilbereich		6	
			Schiefe		1,951	,491
			Kurtosis		4,131	,953
			Mittelwert		3,85	1,271
			95% Konfidenzintervall des Mit-	Untergrenze	1,23	
			telwerts	Obergrenze	6,46	
	2		5% getrimmtes Mittel		2,97	
			Median		,00	
			Varianz		41,975	
			Standardabweichung		6,479	
			Minimum		0	
			Maximum		25	
			Spannweite		25	
			Interquartilbereich		5	
			Schiefe		2,044	,456
			Kurtosis		3,955	,887
			Mittelwert		14,00	4,385
			95% Konfidenzintervall des Mit-	Untergrenze	4,75	
	3		telwerts	Obergrenze	23,25	
			5% getrimmtes Mittel		12,17	
			Median		11,00	
			Varianz		346,118	
			Standardabweichung		18,604	
			Minimum		0	
			Maximum		61	
			Spannweite		61	
			Interquartilbereich		18	
			Schiefe		1,966	,536
			Kurtosis		3,388	1,038
			Mittelwert		19,95	4,610
RE-IX	1		95% Konfidenzintervall des Mit-	Untergrenze	10,37	
			telwerts	Obergrenze	29,54	
			5% getrimmtes Mittel		18,40	
			Median		14,00	

		Varianz		467,474	
		Standardabweichung		21,621	
		Minimum		0	
		Maximum		69	
		Spannweite		69	
		Interquartilbereich		44	
		Schiefe		,917	,491
		Kurtosis		-,366	,953
		Mittelwert		15,65	4,214
		95% Konfidenzintervall des Mittelwerts	Untergrenze	6,98	
			Obergrenze	24,33	
		5% getrimmtes Mittel		13,33	
		Median		8,00	
		Varianz		461,675	
	2	Standardabweichung		21,487	
		Minimum		0	
		Maximum		78	
		Spannweite		78	
		Interquartilbereich		21	
		Schiefe		1,562	,456
		Kurtosis		1,792	,887
		Mittelwert		2,33	1,633
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-1,11	
			Obergrenze	5,78	
		5% getrimmtes Mittel		1,04	
		Median		,00	
		Varianz		48,000	
	3	Standardabweichung		6,928	
		Minimum		0	
		Maximum		28	
		Spannweite		28	
		Interquartilbereich		0	
		Schiefe		3,456	,536
		Kurtosis		12,398	1,038
		Mittelwert		,45	,455
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,49	
			Obergrenze	1,40	
		5% getrimmtes Mittel		,00	
		Median		,00	
		Varianz		4,545	
	1	Standardabweichung		2,132	
		Minimum		0	
		Maximum		10	
		Spannweite		10	
		Interquartilbereich		0	
		Schiefe		4,690	,491
		Kurtosis		22,000	,953
		Mittelwert		5,88	3,285
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,88	
			Obergrenze	12,65	
		5% getrimmtes Mittel		2,69	
		Median		,00	
		Varianz		280,586	
RE-LINK	2	Standardabweichung		16,751	
		Minimum		0	
		Maximum		82	
		Spannweite		82	
		Interquartilbereich		4	
		Schiefe		4,124	,456
		Kurtosis		18,439	,887
		Mittelwert		,17	,167
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,18	
			Obergrenze	,52	
		5% getrimmtes Mittel		,02	
		Median		,00	
		Varianz		,500	
	3	Standardabweichung		,707	
		Minimum		0	
		Maximum		3	
		Spannweite		3	
		Interquartilbereich		0	
		Schiefe		4,243	,536
		Kurtosis		18,000	1,038
		Mittelwert		1,18	1,182
RE-IG	1	95% Konfidenzintervall des Mittelwerts	Untergrenze	-1,28	
			Obergrenze	3,64	
		5% getrimmtes Mittel		,00	

SC-LI	3	Median		,00	
		Varianz		30,727	
		Standardabweichung		5,543	
		Minimum		0	
		Maximum		26	
		Spannweite		26	
		Interquartilbereich		0	
		Schiefe		4,690	,491
		Kurtosis		22,000	,953
		Mittelwert		,78	,558
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,40	
			Obergrenze	1,95	
		5% getrimmtes Mittel		,36	
		Median		,00	
		Varianz		5,595	
		Standardabweichung		2,365	
		Minimum		0	
		Maximum		9	
		Spannweite		9	
		Interquartilbereich		0	
		Schiefe		3,118	,536
		Kurtosis		9,494	1,038
		Mittelwert		1,09	,569
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,09	
			Obergrenze	2,28	
		5% getrimmtes Mittel		,67	
		Median		,00	
		Varianz		7,134	
		Standardabweichung		2,671	
		Minimum		0	
		Maximum		10	
		Spannweite		10	
		Interquartilbereich		1	
SC-LI	1	Schiefe		2,796	,491
		Kurtosis		7,238	,953
		Mittelwert		2,50	1,137
		95% Konfidenzintervall des Mittelwerts	Untergrenze	,16	
			Obergrenze	4,84	
		5% getrimmtes Mittel		1,73	
		Median		,00	
		Varianz		33,620	
		Standardabweichung		5,798	
		Minimum		0	
		Maximum		20	
		Spannweite		20	
		Interquartilbereich		0	
		Schiefe		2,177	,456
		Kurtosis		3,425	,887
		Mittelwert		4,28	1,820
		95% Konfidenzintervall des Mittelwerts	Untergrenze	,44	
			Obergrenze	8,12	
		5% getrimmtes Mittel		3,14	
		Median		,00	
		Varianz		59,624	
		Standardabweichung		7,722	
		Minimum		0	
		Maximum		29	
		Spannweite		29	
		Interquartilbereich		7	
		Schiefe		2,282	,536
		Kurtosis		5,603	1,038
		Mittelwert		27,64	4,342
		95% Konfidenzintervall des Mittelwerts	Untergrenze	18,61	
			Obergrenze	36,67	
		5% getrimmtes Mittel		26,83	
SC-TX	3	Median		,00	
		Varianz		414,719	
		Standardabweichung		20,365	
		Minimum		0	
		Maximum		70	
		Spannweite		70	
		Interquartilbereich		35	
		Schiefe		,745	,491
		Kurtosis		-,360	,953
		Mittelwert		17,54	3,882
		95% Konfidenzintervall des Mittelwerts	Untergrenze	9,54	
			Obergrenze	25,53	
		5% getrimmtes Mittel		26,83	
		Median		22,50	
		Varianz		414,719	
		Standardabweichung		20,365	
		Minimum		0	
		Maximum		70	
		Spannweite		70	
		Interquartilbereich		35	
		Schiefe		,745	,491
		Kurtosis		-,360	,953
		Mittelwert		17,54	3,882
		95% Konfidenzintervall des Mittelwerts	Untergrenze	9,54	
			Obergrenze	25,53	
		5% getrimmtes Mittel		26,83	
		Median		22,50	
		Varianz		414,719	
		Standardabweichung		20,365	
		Minimum		0	
		Maximum		70	
		Spannweite		70	
		Interquartilbereich		35	
		Schiefe		,745	,491
		Kurtosis		-,360	,953
		Mittelwert		17,54	3,882
		95% Konfidenzintervall des Mittelwerts	Untergrenze	9,54	
			Obergrenze	25,53	



SC-IN	3	5% getrimmtes Mittel		15,97	
		Median		10,00	
		Varianz		391,778	
		Standardabweichung		19,793	
		Minimum		0	
		Maximum		67	
		Spannweite		67	
		Interquartilbereich		34	
		Schiefe		1,012	,456
		Kurtosis		-,002	,887
		Mittelwert		32,94	10,201
		95% Konfidenzintervall des Mit-	Untergrenze	11,42	
		telwerts	Obergrenze	54,47	
		5% getrimmtes Mittel		29,55	
		Median		21,50	
	1	Varianz		1873,232	
		Standardabweichung		43,281	
		Minimum		0	
		Maximum		127	
		Spannweite		127	
		Interquartilbereich		39	
		Schiefe		1,443	,536
		Kurtosis		,745	1,038
		Mittelwert		1,23	,411
		95% Konfidenzintervall des Mit-	Untergrenze	,37	
		telwerts	Obergrenze	2,08	
		5% getrimmtes Mittel		1,04	
		Median		,00	
		Varianz		3,708	
		2	Standardabweichung		1,926
Minimum			0		
Maximum			6		
Spannweite			6		
Interquartilbereich			2		
Schiefe			1,450	,491	
Kurtosis			,979	,953	
Mittelwert			1,19	,471	
95% Konfidenzintervall des Mit-	Untergrenze		,22		
telwerts	Obergrenze		2,16		
5% getrimmtes Mittel			,82		
Median			,00		
Varianz			5,762		
3	Standardabweichung			2,400	
	Minimum			0	
	Maximum		10		
	Spannweite		10		
	Interquartilbereich		2		
	Schiefe		2,518	,456	
	Kurtosis		6,844	,887	
	Mittelwert		6,94	2,367	
	95% Konfidenzintervall des Mit-	Untergrenze	1,95		
	telwerts	Obergrenze	11,94		
	5% getrimmtes Mittel		6,10		
	Median		2,00		
	Varianz		100,879		
	1	Standardabweichung		10,044	
		Minimum		0	
Maximum			29		
Spannweite			29		
Interquartilbereich			11		
Schiefe			1,442	,536	
Kurtosis			,812	1,038	
Mittelwert			5,45	1,598	
95% Konfidenzintervall des Mit-		Untergrenze	2,13		
telwerts		Obergrenze	8,78		
5% getrimmtes Mittel			4,90		
Median			,00		
Varianz			56,165		
2		Standardabweichung		7,494	
		Minimum		0	
	Maximum		21		
	Spannweite		21		
	Interquartilbereich		12		
	Schiefe		1,037	,491	
	Kurtosis		-,452	,953	
	Mittelwert		10,00	3,024	
	95% Konfidenzintervall des Mit-	Untergrenze	3,77		

SC-IX	3	telwerts	Obergrenze	16,23	
		5% getrimmtes Mittel		7,70	
		Median		4,00	
		Varianz		237,760	
		Standardabweichung		15,419	
		Minimum		0	
		Maximum		69	
		Spannweite		69	
		Interquartilbereich		16	
		Schiefe		2,586	,456
		Kurtosis		8,089	,887
		Mittelwert		7,17	2,782
		95% Konfidenzintervall des Mit-	Untergrenze	1,30	
		telwerts	Obergrenze	13,04	
		5% getrimmtes Mittel		5,35	
		Median		3,00	
		Varianz		139,324	
		Standardabweichung		11,804	
		Minimum		0	
		Maximum		47	
		Spannweite		47	
		Interquartilbereich		7	
		Schiefe		2,628	,536
		Kurtosis		7,637	1,038
		Mittelwert		5,82	1,417
		95% Konfidenzintervall des Mit-	Untergrenze	2,87	
		telwerts	Obergrenze	8,76	
		5% getrimmtes Mittel		5,07	
		Median		4,50	
		Varianz		44,156	
		Standardabweichung		6,645	
SC-IX	1	Minimum		0	
		Maximum		26	
		Spannweite		26	
		Interquartilbereich		10	
		Schiefe		1,514	,491
		Kurtosis		2,790	,953
		Mittelwert		4,46	1,505
		95% Konfidenzintervall des Mit-	Untergrenze	1,36	
		telwerts	Obergrenze	7,56	
		5% getrimmtes Mittel		3,45	
		Median		,00	
		Varianz		58,898	
		Standardabweichung		7,675	
		Minimum		0	
		Maximum		29	
		Spannweite		29	
		Interquartilbereich		6	
		Schiefe		2,026	,456
		Kurtosis		3,611	,887
		Mittelwert		2,94	,730
		95% Konfidenzintervall des Mit-	Untergrenze	1,40	
		telwerts	Obergrenze	4,48	
		5% getrimmtes Mittel		2,83	
		Median		1,50	
		Varianz		9,585	
		Standardabweichung		3,096	
		Minimum		0	
		Maximum		8	
		Spannweite		8	
		Interquartilbereich		6	
		Schiefe		,487	,536
		Kurtosis		-1,649	1,038
		Mittelwert		1,91	1,143
		95% Konfidenzintervall des Mit-	Untergrenze	-,47	
		telwerts	Obergrenze	4,29	
SC-LINK	1	5% getrimmtes Mittel		1,12	
		Median		,00	
		Varianz		28,753	
		Standardabweichung		5,362	
		Minimum		0	
		Maximum		18	
		Spannweite		18	
		Interquartilbereich		0	
		Schiefe		2,809	,491
		Kurtosis		6,842	,953
		Mittelwert		2,38	1,023

EX-PI	3	95% Konfidenzintervall des Mittelwerts	Untergrenze	,28	
			Obergrenze	4,49	
		5% getrimmtes Mittel		1,52	
		Median		,00	
		Varianz		27,206	
		Standardabweichung		5,216	
		Minimum		0	
		Maximum		24	
		Spannweite		24	
		Interquartilbereich		5	
		Schiefe		3,158	,456
		Kurtosis		11,821	,887
		Mittelwert		4,17	1,344
		95% Konfidenzintervall des Mittelwerts	Untergrenze	1,33	
			Obergrenze	7,00	
		5% getrimmtes Mittel		3,69	
		Median		1,00	
		Varianz		32,500	
		Standardabweichung		5,701	
		Minimum		0	
		Maximum		17	
		Spannweite		17	
		Interquartilbereich		8	
		Schiefe		1,190	,536
		Kurtosis		,027	1,038
		Mittelwert		2,64	,746
EX-PI	1	95% Konfidenzintervall des Mittelwerts	Untergrenze	1,09	
			Obergrenze	4,19	
		5% getrimmtes Mittel		2,37	
		Median		1,00	
		Varianz		12,242	
		Standardabweichung		3,499	
		Minimum		0	
		Maximum		10	
		Spannweite		10	
		Interquartilbereich		5	
		Schiefe		1,142	,491
		Kurtosis		-,041	,953
		Mittelwert		30,46	4,971
		95% Konfidenzintervall des Mittelwerts	Untergrenze	20,22	
			Obergrenze	40,70	
		5% getrimmtes Mittel		28,38	
		Median		26,00	
		Varianz		642,418	
		Standardabweichung		25,346	
		Minimum		0	
		Maximum		98	
		Spannweite		98	
		Interquartilbereich		20	
		Schiefe		1,542	,456
		Kurtosis		2,495	,887
		Mittelwert		13,61	3,399
EX-PI	2	95% Konfidenzintervall des Mittelwerts	Untergrenze	6,44	
			Obergrenze	20,78	
		5% getrimmtes Mittel		12,18	
		Median		9,50	
		Varianz		208,016	
		Standardabweichung		14,423	
		Minimum		0	
		Maximum		53	
		Spannweite		53	
		Interquartilbereich		20	
		Schiefe		1,241	,536
		Kurtosis		1,706	1,038
		Mittelwert		1,18	1,044
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,99	
			Obergrenze	3,35	
		5% getrimmtes Mittel		,14	
		Median		,00	
		Varianz		23,965	
SC-IB	1	Standardabweichung		4,895	
		Minimum		0	
		Maximum		23	
		Spannweite		23	
		Interquartilbereich		0	
		Schiefe		4,623	,491
		Kurtosis		21,541	,953

EX-IB	2	Mittelwert		,77	,769
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,82	
			Obergrenze	2,35	
		5% getrimmtes Mittel		,00	
		Median		,00	
		Varianz		15,385	
		Standardabweichung		3,922	
		Minimum		0	
		Maximum		20	
		Spannweite		20	
		Interquartilbereich		0	
		Schiefe		5,099	,456
		Kurtosis		26,000	,887
	3	Mittelwert		1,78	,650
		95% Konfidenzintervall des Mittelwerts	Untergrenze	,41	
			Obergrenze	3,15	
		5% getrimmtes Mittel		1,48	
		Median		,00	
		Varianz		7,595	
		Standardabweichung		2,756	
		Minimum		0	
		Maximum		9	
		Spannweite		9	
		Interquartilbereich		4	
		Schiefe		1,493	,536
		Kurtosis		1,299	1,038
EX-IG	2	Mittelwert		,23	,231
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,24	
			Obergrenze	,71	
		5% getrimmtes Mittel		,00	
		Median		,00	
		Varianz		1,385	
		Standardabweichung		1,177	
		Minimum		0	
		Maximum		6	
		Spannweite		6	
		Interquartilbereich		0	
		Schiefe		5,099	,456
		Kurtosis		26,000	,887
	1	Mittelwert		1,82	,772
		95% Konfidenzintervall des Mittelwerts	Untergrenze	,21	
			Obergrenze	3,42	
		5% getrimmtes Mittel		1,27	
		Median		,00	
		Varianz		13,108	
		Standardabweichung		3,621	
		Minimum		0	
		Maximum		14	
		Spannweite		14	
		Interquartilbereich		3	
		Schiefe		2,387	,491
		Kurtosis		5,771	,953
EX-IG	2	Mittelwert		,31	,182
		95% Konfidenzintervall des Mittelwerts	Untergrenze	-,07	
			Obergrenze	,68	
		5% getrimmtes Mittel		,15	
		Median		,00	
		Varianz		,862	
		Standardabweichung		,928	
		Minimum		0	
		Maximum		4	
		Spannweite		4	
		Interquartilbereich		0	
		Schiefe		3,217	,456
		Kurtosis		10,480	,887
	3	Mittelwert		2,33	,886
		95% Konfidenzintervall des Mittelwerts	Untergrenze	,46	
			Obergrenze	4,20	
		5% getrimmtes Mittel		1,98	
		Median		,00	
		Varianz		14,118	
		Standardabweichung		3,757	
		Minimum		0	
		Maximum		11	
		Spannweite		11	
		Interquartilbereich		4	
		Schiefe		1,616	,536

NA	1	Kurtosis		1,484	1,038
		Mittelwert		32,23	3,345
		95% Konfidenzintervall des Mittelwerts	Untergrenze	25,27	
			Obergrenze	39,18	
		5% getrimmtes Mittel		31,67	
		Median		35,00	
		Varianz		246,089	
		Standardabweichung		15,687	
		Minimum		8	
		Maximum		68	
	2	Spannweite		60	
		Interquartilbereich		24	
		Schiefe		,245	,491
		Kurtosis		-,340	,953
		Mittelwert		27,58	3,566
		95% Konfidenzintervall des Mittelwerts	Untergrenze	20,23	
			Obergrenze	34,92	
		5% getrimmtes Mittel		26,39	
		Median		21,50	
		Varianz		330,574	
	3	Standardabweichung		18,182	
		Minimum		2	
		Maximum		75	
		Spannweite		73	
		Interquartilbereich		25	
		Schiefe		1,149	,456
		Kurtosis		,974	,887
		Mittelwert		28,33	6,619
		95% Konfidenzintervall des Mittelwerts	Untergrenze	14,37	
			Obergrenze	42,30	
SC	1	5% getrimmtes Mittel		25,70	
		Median		21,50	
		Varianz		788,706	
		Standardabweichung		28,084	
		Minimum		6	
		Maximum		98	
		Spannweite		92	
		Interquartilbereich		20	
		Schiefe		1,917	,536
		Kurtosis		2,968	1,038
	2	Mittelwert		4,318	1,4194
		95% Konfidenzintervall des Mittelwerts	Untergrenze	1,366	
			Obergrenze	7,270	
		5% getrimmtes Mittel		3,631	
		Median		1,000	
		Varianz		44,323	
		Standardabweichung		6,6575	
		Minimum		,0	
		Maximum		21,0	
		Spannweite		21,0	
	3	Interquartilbereich		6,0	
		Schiefe		1,780	,491
		Kurtosis		2,251	,953
		Mittelwert		6,923	1,7715
		95% Konfidenzintervall des Mittelwerts	Untergrenze	3,275	
			Obergrenze	10,572	
		5% getrimmtes Mittel		6,000	
		Median		2,500	
		Varianz		81,594	
		Standardabweichung		9,0329	
	2	Minimum		,0	
		Maximum		33,0	
		Spannweite		33,0	
		Interquartilbereich		13,3	
		Schiefe		1,290	,456
		Kurtosis		1,157	,887
		Mittelwert		7,222	2,0904
		95% Konfidenzintervall des Mittelwerts	Untergrenze	2,812	
			Obergrenze	11,633	
		5% getrimmtes Mittel		6,469	

Median	5,500	
Varianz	78,654	
Standardabweichung	8,8687	
Minimum	,0	
Maximum	28,0	
Spannweite	28,0	
Interquartilbereich	8,3	
Schiefe	1,599	,536
Kurtosis	1,837	1,038

a. RE-IG ist bei V1 = 2 konstant und wurde weggelassen.

b. EX-IB ist bei V1 = 1 konstant und wurde weggelassen.

c. EX-IB ist bei V1 = 3 konstant und wurde weggelassen.

### 8.3. Test auf Normalverteilung

Tabelle 20-26 Test auf Normalverteilung relative Häufigkeiten look up

Tests auf Normalverteilung <sup>b</sup>						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
RE-LI	,358	31	,000	,540	31	,000
RE-IN	,223	31	,000	,822	31	,000
RE-TX	,238	31	,000	,838	31	,000
RE-HL	,379	31	,000	,611	31	,000
RE-IX	,273	31	,000	,735	31	,000
RE-LINK	,539	31	,000	,176	31	,000
RE-IG	,539	31	,000	,176	31	,000
SC-LI	,458	31	,000	,456	31	,000
SC-TX	,154	31	,060	,932	31	,048
SC-IN	,375	31	,000	,676	31	,000
SC-HL	,259	31	,000	,697	31	,000
SC-IX	,236	31	,000	,778	31	,000
SC-LINK	,436	31	,000	,491	31	,000
EX-PI	,261	31	,000	,773	31	,000
SC-IB	,435	31	,000	,330	31	,000
EX-IG	,371	31	,000	,357	31	,000

a. Signifikanzkorrektur nach Lilliefors

b. EX-IB ist konstant und wurde weggelassen.

Tabelle 20-27 Test auf Normalverteilung relative Häufigkeiten casual leisure

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
RE-LI	,375	34	,000	,408	34	,000
RE-IN	,205	34	,001	,857	34	,000
RE-TX	,183	34	,005	,889	34	,002
RE-HL	,286	34	,000	,590	34	,000
RE-IX	,507	34	,000	,266	34	,000
RE-LINK	,539	34	,000	,165	34	,000
RE-IG	,535	34	,000	,256	34	,000
SC-LI	,411	34	,000	,443	34	,000
SC-TX	,183	34	,006	,828	34	,000
SC-IN	,257	34	,000	,756	34	,000
SC-HL	,278	34	,000	,752	34	,000
SC-IX	,310	34	,000	,712	34	,000
SC-LINK	,311	34	,000	,506	34	,000
EX-PI	,224	34	,000	,758	34	,000
SC-IB	,326	34	,000	,524	34	,000
EX-IB	,529	34	,000	,326	34	,000
EX-IG	,357	34	,000	,394	34	,000

a. Signifikanzkorrektur nach Lilliefors

Tabelle 20-28 Test auf Normalverteilung, relative Häufigkeiten learn

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
BE-BI	,165	42	,006	,860	42	,000
BE-IB	,399	42	,000	,479	42	,000
BE-IG	,289	42	,000	,636	42	,000
LE-AA	,174	42	,003	,914	42	,004
LE-EL	,256	42	,000	,688	42	,000
LE-IB	,501	42	,000	,261	42	,000
LE-LI	,327	42	,000	,506	42	,000
LE-LK	,521	42	,000	,393	42	,000
LE-UES	,287	42	,000	,620	42	,000
QL-AA	,231	42	,000	,766	42	,000
QL-EL	,288	42	,000	,729	42	,000
QL-IB	,336	42	,000	,504	42	,000
QL-IG	,435	42	,000	,444	42	,000
QL-LI	,226	42	,000	,853	42	,000
QL-LK	,410	42	,000	,442	42	,000
QL-UES	,206	42	,000	,800	42	,000

a. Signifikanzkorrektur nach Lilliefors

Tabelle 20-29 Test auf Normalverteilung über alle Gruppen, relative Häufigkeiten

Tests auf Normalverteilung						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
RE-LI	,363	101	,000	,474	101	,000
RE-IN	,187	101	,000	,810	101	,000
RE-TX	,131	101	,000	,905	101	,000
RE-HL	,327	101	,000	,493	101	,000
RE-IX	,369	101	,000	,588	101	,000
RE-LINK	,485	101	,000	,185	101	,000
RE-IG	,527	101	,000	,130	101	,000
SC-LI	,422	101	,000	,448	101	,000
SC-TX	,172	101	,000	,849	101	,000
SC-IN	,307	101	,000	,570	101	,000
SC-HL	,240	101	,000	,738	101	,000
SC-IX	,250	101	,000	,707	101	,000
SC-LINK	,315	101	,000	,550	101	,000
EX-PI	,234	101	,000	,731	101	,000
SC-IB	,401	101	,000	,303	101	,000
EX-IB	,539	101	,000	,195	101	,000
EX-IG	,394	101	,000	,284	101	,000
Summe	,173	101	,000	,931	101	,000
Valence	,201	101	,000	,900	101	,000
Arousal	,254	101	,000	,823	101	,000
Geschlecht (1=w, 2=m)	,429	101	,000	,591	101	,000
Alter	,158	101	,000	,924	101	,000
WP-Kenntnis	,350	101	,000	,797	101	,000
Vorkenntnis Thema	,233	101	,000	,842	101	,000

a. Signifikanzkorrektur nach Lilliefors

## 8.4. Test auf Varianzhomogenität

Tabelle 20-30 Test auf Homogenität der Varianzen alle drei TT, relative Häufigkeiten

Test der Homogenität der Varianzen				
	Levene-Statistik	df1	df2	Signifikanz
RE-LI	4,380	2	98	,015
RE-IN	26,506	2	98	,000
RE-TX	12,181	2	98	,000
RE-HL	9,921	2	98	,000
RE-IX	12,236	2	98	,000
RE-LINK	7,536	2	98	,001
RE-IG	3,187	2	98	,046
SC-LI	7,761	2	98	,001
SC-TX	2,499	2	98	,087
SC-IN	16,699	2	98	,000
SC-HL	1,477	2	98	,233
SC-IX	,130	2	98	,878
SC-LINK	6,332	2	98	,003
EX-PI	17,281	2	98	,000
SC-IB	4,047	2	98	,020
EX-IB	10,094	2	98	,000
EX-IG	4,522	2	98	,013

## 8.5. Relative Werte

Tabelle 20-31 Relative Werte Studie zwei ( $n_{\text{Nutzeraktion}}/n_{\text{Summe}}$ )

Relativ CL	Relativ LE	Relativ LU
0,026767789	0,019177391	0,095806452
0,169529333	0,127554883	0,056129032
0,428507696	0,540878122	0,325806451
0,056212358	0,012616705	0,022903226
0,009368726	0,051349987	0,141612903
0,000669195	0,019303558	0,003225806
0,003122909	3,28034E-10	0,008387097
0,017175998	0,008200858	0,007741935
0,132277493	0,057532173	0,196129032
0,027883114	0,003911178	0,008709677
0,028775374	0,032803432	0,038709677
0,01182244	0,014635377	0,041290323
0,016729868	0,007822357	0,013548387
0,054650903	0,0999243	0,018709677
0,007138077	0,002523341	0,008387097
4,23823E-10	0,000757002	7,09677E-10
0,009368726	0,001009336	0,012903226



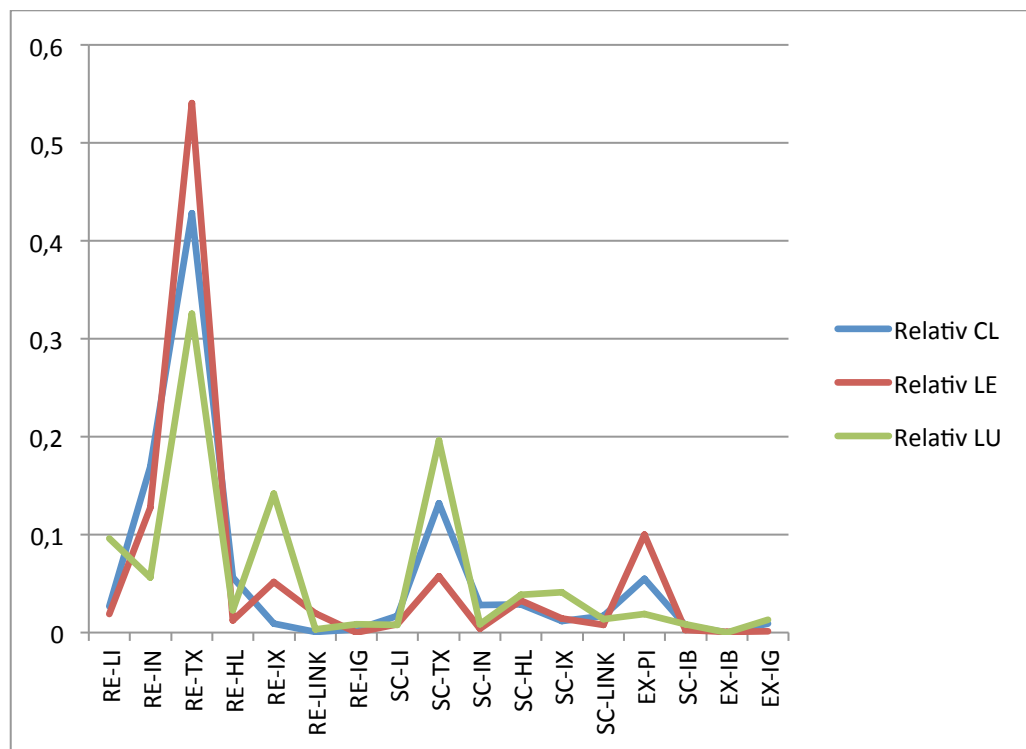


Abbildung 20-1 relative Häufigkeiten Taskweise ( $n_{\text{Nutzeraktion}}/n_{\text{Summe}}$ )

## 8.6. Themenvergleich L1 und L2

Tabelle 20-32 Vergleich der Look-up Tasks Thema 1 vs. Thema 2, t-Test, relative Häufigkeiten

## Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
RE-LI	Varianzen sind gleich	2,253	,153	,834	16	,417	9,778	11,726	-15,081	34,636
	Varianzen sind nicht gleich			,834	10,349	,423	9,778	11,726	-16,231	35,787
RE-IN	Varianzen sind gleich	2,130	,164	3,503	16	,003	12,111	3,457	4,783	19,440
	Varianzen sind nicht gleich			3,503	11,559	,005	12,111	3,457	4,547	19,675
RE-TX	Varianzen sind gleich	1,071	,316	-,683	16	,505	-15,444	22,625	-63,407	32,518
	Varianzen sind nicht gleich			-,683	15,135	,505	-15,444	22,625	-63,631	32,742
RE-HL	Varianzen sind gleich	2,688	,121	,626	16	,540	1,667	2,661	-3,974	7,307
	Varianzen sind nicht gleich			,626	12,562	,542	1,667	2,661	-4,102	7,436
RE-IX	Varianzen sind gleich	9,890	,006	-	16	,178	-14,333	10,176	-35,906	7,239
	Varianzen sind nicht gleich			-	12,196	,184	-14,333	10,176	-36,466	7,799
RE-LINK	Varianzen sind gleich	5,224	,036	1,000	16	,332	1,111	1,111	-1,244	3,467
	Varianzen sind nicht gleich			1,000	8,000	,347	1,111	1,111	-1,451	3,673
RE-IG	Varianzen sind gleich	5,224	,036	1,000	16	,332	2,889	2,889	-3,235	9,013
	Varianzen sind nicht gleich			1,000	8,000	,347	2,889	2,889	-3,773	9,551
SC-LI	Varianzen sind gleich	14,615	,001	1,908	16	,075	2,444	1,281	-,272	5,161
	Varianzen sind nicht gleich			1,908	8,000	,093	2,444	1,281	-,510	5,399
SC-TX	Varianzen sind gleich	10,182	,006	-	16	,258	-11,778	10,039	-33,059	9,503
	Varianzen sind nicht gleich			-	11,221	,265	-11,778	10,039	-33,820	10,264
SC-IN	Varianzen sind gleich	3,184	,093	1,012	16	,327	-,889	,878	-2,751	,973
	Varianzen sind nicht gleich			1,012	11,931	,332	-,889	,878	-2,804	1,026
SC-HL	Varianzen sind gleich	13,331	,002	2,619	16	,019	8,222	3,139	1,567	14,877
	Varianzen sind nicht gleich			2,619	10,319	,025	8,222	3,139	1,257	15,188
SC-IX	Varianzen sind gleich	,163	,692	-,365	16	,720	-1,222	3,345	-8,313	5,869
	Varianzen sind nicht gleich			-,365	14,590	,720	-1,222	3,345	-8,369	5,925
SC-LINK	Varianzen sind gleich	2,090	,168	,632	16	,536	1,333	2,108	-3,136	5,802
	Varianzen sind nicht gleich			,632	9,756	,542	1,333	2,108	-3,380	6,047
EX-PI	Varianzen sind gleich	14,405	,002	1,384	16	,185	2,111	1,526	-1,123	5,345
	Varianzen sind nicht gleich			1,384	10,549	,195	2,111	1,526	-1,264	5,486
SC-IB	Varianzen sind gleich	4,077	,061	,866	16	,399	2,222	2,566	-3,218	7,663
	Varianzen sind nicht gleich			,866	8,136	,411	2,222	2,566	-3,679	8,123
EX-IG	Varianzen sind gleich	14,415	,002	2,414	16	,028	3,889	1,611	,473	7,304
	Varianzen sind nicht gleich			2,414	8,000	,042	3,889	1,611	,174	7,604
NA	Varianzen sind gleich	,569	,462	-,645	16	,528	-5,000	7,748	-21,424	11,424
	Varianzen sind nicht gleich			-,645	14,288	,529	-5,000	7,748	-21,585	11,585
SC	Varianzen sind gleich	,408	,532	,152	16	,881	,444	2,928	-5,762	6,651
	Varianzen sind nicht gleich			,152	14,650	,881	,444	2,928	-5,809	6,698

## 8.7. Vergleich der Gruppen emotional positiv und negativ

Tabelle 20-33 t-Test Gruppe arousal positiv und arousal negativ, Trennwert 4, Wahrscheinlichkeiten der Videolabelkombinationen

		Test bei unabhängigen Stichproben								
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
RE-LI	Varianzen sind gleich	1,773	,188	-,494	64	,623	-2,398	4,851	-12,088	7,293
	Varianzen sind nicht gleich			-,368	17,911	,717	-2,398	6,513	-16,086	11,291
RE-IN	Varianzen sind gleich	1,475	,229	1,314	64	,193	10,850	8,255	-5,641	27,341
	Varianzen sind nicht gleich			1,583	36,744	,122	10,850	6,853	-3,038	24,738
RE-TX	Varianzen sind gleich	,162	,689	2,712	64	,009	66,315	24,452	17,467	115,163
	Varianzen sind nicht gleich			2,718	25,433	,012	66,315	24,398	16,110	116,520
RE-HL	Varianzen sind gleich	3,447	,068	1,652	64	,103	5,490	3,322	-1,147	12,127
	Varianzen sind nicht gleich			2,408	58,586	,019	5,490	2,280	,927	10,053
RE-IX	Varianzen sind gleich	,804	,373	1,057	64	,295	-6,000	5,679	-17,345	5,345
	Varianzen sind nicht gleich			-,996	23,174	,330	-6,000	6,024	-18,456	6,456
RE-LINK	Varianzen sind gleich	,552	,460	,402	64	,689	1,257	3,127	-4,990	7,505
	Varianzen sind nicht gleich			,616	63,078	,540	1,257	2,042	-2,824	5,339
SC-LI	Varianzen sind gleich	8,766	,004	1,543	64	,128	2,495	1,617	-,736	5,726
	Varianzen sind nicht gleich			2,423	63,975	,018	2,495	1,030	,438	4,552
SC-TX	Varianzen sind gleich	,050	,824	1,183	64	,241	-9,677	8,183	-26,025	6,670
	Varianzen sind nicht gleich			1,373	33,761	,179	-9,677	7,047	-24,003	4,648
SC-IN	Varianzen sind gleich	,454	,503	,729	64	,469	1,268	1,740	-2,208	4,743
	Varianzen sind nicht gleich			,815	31,160	,421	1,268	1,555	-1,904	4,439
SC-HL	Varianzen sind gleich	3,541	,064	1,482	64	,143	-5,165	3,485	-12,127	1,797
	Varianzen sind nicht gleich			1,123	18,175	,276	-5,165	4,599	-14,820	4,490
SC-IX	Varianzen sind gleich	1,909	,172	1,453	64	,151	-2,640	1,817	-6,269	,989
	Varianzen sind nicht gleich			1,253	20,663	,224	-2,640	2,108	-7,028	1,748
EX-PI	Varianzen sind gleich	4,449	,039	2,419	64	,018	14,228	5,883	2,475	25,980
	Varianzen sind nicht gleich			3,453	56,111	,001	14,228	4,121	5,973	22,482
SC-IB	Varianzen sind gleich	,247	,621	-,150	64	,881	-,173	1,151	-2,472	2,127
	Varianzen sind nicht gleich			-,128	20,404	,900	-,173	1,350	-2,985	2,640
EX-IB	Varianzen sind gleich	14,842	,000	1,798	64	,077	-,375	,209	-,792	,042
	Varianzen sind nicht gleich			1,000	15,000	,333	-,375	,375	-1,174	,424
EX-IG	Varianzen sind gleich	2,139	,149	-,588	64	,559	-,510	,868	-2,244	1,224
	Varianzen sind nicht gleich			-,474	19,235	,641	-,510	1,077	-2,762	1,742
NA	Varianzen sind gleich	,003	,959	1,036	64	,304	-6,078	5,864	-17,792	5,637
	Varianzen sind nicht gleich			1,077	27,024	,291	-6,078	5,645	-17,661	5,506
SC	Varianzen sind gleich	,593	,444	-,201	64	,841	-,480	2,386	-5,246	4,286
	Varianzen sind nicht gleich			-,186	22,523	,854	-,480	2,583	-5,830	4,870

Tabelle 20-34 t-Test Gruppe valence positiv und valence negativ, Trennwert 4, relative Häufigkeiten der Videolabelkombinationen

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
RE-LI	Varianzen sind gleich	,756	,394	,160	21	,874	1,603	10,007	-19,207	22,413
	Varianzen sind nicht gleich			,138	10,349	,893	1,603	11,580	-24,082	27,288
RE-IN	Varianzen sind gleich	,525	,477	-,486	21	,632	-2,778	5,720	-14,672	9,117
	Varianzen sind nicht gleich			-,531	20,953	,601	-2,778	5,230	-13,656	8,101
RE-TX	Varianzen sind gleich	,002	,966	,119	21	,906	2,921	24,464	-47,954	53,795
	Varianzen sind nicht gleich			,125	19,514	,902	2,921	23,436	-46,044	51,886
RE-HL	Varianzen sind gleich	13,227	,002	-	21	,016	-5,071	1,940	-9,106	-1,037
	Varianzen sind nicht gleich			2,614	13,000	,006	-5,071	1,542	-8,403	-1,739
RE-IX	Varianzen sind gleich	16,423	,001	-	21	,007	-23,516	7,917	-39,981	-7,051
	Varianzen sind nicht gleich			2,970	16,932	,002	-23,516	6,585	-37,414	-9,618
RE-LINK	Varianzen sind gleich	2,967	,100	-,795	21	,435	-,714	,898	-2,583	1,154
	Varianzen sind nicht gleich			-	13,000	,336	-,714	,714	-2,257	,829
SC-LI	Varianzen sind gleich	8,963	,007	-	21	,221	-1,841	1,460	-4,877	1,194
	Varianzen sind nicht gleich			1,262	14,665	,141	-1,841	1,183	-4,367	,685
SC-TX	Varianzen sind gleich	1,509	,233	-,796	21	,435	-7,103	8,925	-25,663	11,457
	Varianzen sind nicht gleich			-,827	19,287	,418	-7,103	8,590	-25,065	10,858
SC-IN	Varianzen sind gleich	,199	,660	,010	21	,992	,008	,823	-1,703	1,719
	Varianzen sind nicht gleich			,009	15,235	,993	,008	,852	-1,806	1,822
SC-HL	Varianzen sind gleich	,520	,479	-,268	21	,791	-,857	3,200	-7,511	5,797
	Varianzen sind nicht gleich			-,283	19,999	,780	-,857	3,032	-7,181	5,467
SC-IX	Varianzen sind gleich	1,927	,180	-,183	21	,856	-,524	2,858	-6,467	5,420
	Varianzen sind nicht gleich			-,208	20,607	,838	-,524	2,524	-5,779	4,732
SC-LINK	Varianzen sind gleich	3,947	,060	-,843	21	,409	-1,905	2,260	-6,604	2,794
	Varianzen sind nicht gleich			-	16,494	,323	-1,905	1,870	-5,859	2,049
EX-PI	Varianzen sind gleich	,055	,817	,124	21	,903	,270	2,177	-4,257	4,797
	Varianzen sind nicht gleich			,132	20,364	,896	,270	2,043	-3,986	4,526
SC-IB	Varianzen sind gleich	2,986	,099	-,904	21	,376	-1,857	2,055	-6,130	2,416
	Varianzen sind nicht gleich			-	13,000	,276	-1,857	1,634	-5,386	1,672
EX-IG	Varianzen sind gleich	3,663	,069	1,137	21	,327	1,524	1,520	-1,637	4,684
	Varianzen sind nicht gleich			1,003	10,953	,397	1,524	1,729	-2,283	5,330
NA	Varianzen sind gleich	,514	,481	-	21	,010	-16,365	5,763	-28,351	-4,380
	Varianzen sind nicht gleich			2,840	19,795	,007	-16,365	5,487	-27,819	-4,912
SC	Varianzen sind gleich	6,955	,015	2,982	21	,277	-3,095	2,771	-8,858	2,668
	Varianzen sind nicht gleich			1,117	17,019	,197	-3,095	2,307	-7,963	1,773

## 8.8. Vergleich der Gruppe viel und wenig Wikipedia -Vorkenntnis

Tabelle 20-35 t-Test Gruppen viel und wenig WP-Kenntnis, Trennwert 2, relative Häufigkeiten der Videolabelkombinationen

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
RE-LI	Varianzen sind gleich	,972	,327	-,014	99	,989	-,063	4,402	-8,797	8,672
	Varianzen sind nicht gleich			-,026	32,397	,979	-,063	2,405	-4,960	4,834
RE-IN	Varianzen sind gleich	3,360	,070	,635	99	,527	6,890	10,851	-14,642	28,421
	Varianzen sind nicht gleich			1,113	29,000	,275	6,890	6,190	-5,771	19,550
RE-TX	Varianzen sind gleich	6,393	,013	,319	99	,751	9,879	30,985	-51,602	71,361
	Varianzen sind nicht gleich			,475	21,079	,640	9,879	20,797	-33,360	53,118
RE-HL	Varianzen sind gleich	1,486	,226	,783	99	,435	2,386	3,046	-3,657	8,429
	Varianzen sind nicht gleich			1,285	25,018	,211	2,386	1,857	-1,439	6,211
RE-IX	Varianzen sind gleich	4,308	,041	,436	99	,664	2,317	5,312	-8,223	12,858
	Varianzen sind nicht gleich			,872	42,404	,388	2,317	2,658	-3,045	7,680
RE-LINK	Varianzen sind gleich	,954	,331	,477	99	,634	1,298	2,718	-4,095	6,691
	Varianzen sind nicht gleich			1,238	98,940	,219	1,298	1,048	-,783	3,378
SC-LI	Varianzen sind gleich	,090	,765	-,428	99	,669	-,900	2,100	-5,067	3,267
	Varianzen sind nicht gleich			-,364	13,006	,722	-,900	2,474	-6,245	4,446
SC-TX	Varianzen sind gleich	,044	,834	-,870	99	,387	-7,370	8,474	-24,183	9,444
	Varianzen sind nicht gleich			-,723	12,895	,483	-7,370	10,194	-29,410	14,671
SC-IN	Varianzen sind gleich	7,531	,007	1,599	99	,113	3,889	2,432	-,937	8,714
	Varianzen sind nicht gleich			4,130	98,999	,000	3,889	,942	2,020	5,757
SC-HL	Varianzen sind gleich	1,371	,244	,996	99	,322	5,139	5,159	-5,099	15,376
	Varianzen sind nicht gleich			1,192	16,166	,251	5,139	4,312	-3,995	14,272
SC-IX	Varianzen sind gleich	11,634	,001	2,228	99	,028	10,914	4,897	1,196	20,632
	Varianzen sind nicht gleich			5,777	98,939	,000	10,914	1,889	7,166	14,662
SC-LINK	Varianzen sind gleich	6,541	,012	1,591	99	,115	5,884	3,699	-1,455	13,223
	Varianzen sind nicht gleich			4,132	98,868	,000	5,884	1,424	3,059	8,709
EX-PI	Varianzen sind gleich	,511	,476	-,041	99	,968	-,308	7,596	-15,380	14,764
	Varianzen sind nicht gleich			-,052	17,301	,959	-,308	5,929	-12,801	12,185
SC-IB	Varianzen sind gleich	1,231	,270	,600	99	,550	1,759	2,930	-4,054	7,573
	Varianzen sind nicht gleich			1,324	61,618	,190	1,759	1,329	-,897	4,415
EX-IB	Varianzen sind gleich	2,316	,131	,728	99	,468	,270	,370	-,465	1,004
	Varianzen sind nicht gleich			1,992	88,000	,049	,270	,135	,001	,539
EX-IG	Varianzen sind gleich	,532	,467	,240	99	,811	,759	3,167	-5,524	7,043
	Varianzen sind nicht gleich			,502	50,252	,618	,759	1,512	-2,277	3,796

## 8.9. Vergleich der Gruppen viel und wenig Themen-Vorkenntnis

Tabelle 20-36 t-Test zw. Gruppen viel und wenig thematische Vorkenntnis, Trennwert 2

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
RE-LI	Varianzen sind gleich	,474	,493	,352	99	,726	2,949	8,385	-13,688	19,586
	Varianzen sind nicht gleich			,884	3,053	,441	2,949	3,336	-7,563	13,461
RE-IN	Varianzen sind gleich	3,291	,073	,860	99	,392	17,765	20,647	-23,202	58,733
	Varianzen sind nicht gleich			4,548	52,371	,000	17,765	3,906	9,929	25,601
RE-TX	Varianzen sind gleich	1,358	,247	-,289	99	,773	-17,051	59,062	-134,243	100,141
	Varianzen sind nicht gleich			-,449	2,326	,692	-17,051	37,964	-160,302	126,200
RE-HL	Varianzen sind gleich	1,625	,205	,788	99	,433	4,571	5,805	-6,946	16,089
	Varianzen sind nicht gleich			4,523	97,000	,000	4,571	1,011	2,566	6,577
RE-IX	Varianzen sind gleich	,957	,330	,047	99	,963	,473	10,134	-19,635	20,581
	Varianzen sind nicht gleich			,103	2,742	,925	,473	4,606	-14,996	15,942
RE-LINK	Varianzen sind gleich	,125	,724	,128	99	,898	,663	5,186	-9,627	10,953
	Varianzen sind nicht gleich			,492	6,497	,639	,663	1,347	-2,573	3,899
RE-IG	Varianzen sind gleich	,957	,330	-,806	99	,422	-1,310	1,625	-4,534	1,915
	Varianzen sind nicht gleich			-,775	2,114	,516	-1,310	1,690	-8,217	5,598
SC-LI	Varianzen sind gleich	7,156	,009	-	99	,024	-8,976	3,904	-16,722	-1,231
	Varianzen sind nicht gleich			2,299	-					
SC-TX	Varianzen sind gleich	7,522	,007	1,013	2,021	,417	-8,976	8,861	-46,723	28,771
	Varianzen sind nicht gleich			-	99	,043	-32,544	15,878	-64,051	-1,038
SC-IN	Varianzen sind gleich	2,381	,126	-,928	2,023	,450	-32,544	35,071	-181,843	116,755
	Varianzen sind nicht gleich			,884	99	,379	4,133	4,676	-5,146	13,411
SC-HL	Varianzen sind gleich	,716	,399	5,076	97,000	,000	4,133	,814	2,517	5,749
	Varianzen sind nicht gleich			-,887	99	,377	-8,731	9,844	-28,263	10,801
SC-IX	Varianzen sind gleich	2,530	,115	-,622	2,058	,596	-8,731	14,033	-67,498	50,035
	Varianzen sind nicht gleich			,932	99	,353	8,881	9,524	-10,017	27,778
SC-LINK	Varianzen sind gleich	1,209	,274	4,029	10,240	,002	8,881	2,204	3,985	13,777
	Varianzen sind nicht gleich			,654	99	,515	4,656	7,124	-9,479	18,792
EX-PI	Varianzen sind gleich	,009	,926	2,557	6,849	,038	4,656	1,821	,332	8,981
	Varianzen sind nicht gleich			-,679	99	,498	-9,813	14,444	-38,472	18,846
SC-IB	Varianzen sind gleich	,284	,595	-,736	2,148	,534	-9,813	13,329	-63,546	43,920
	Varianzen sind nicht gleich			,240	99	,811	1,340	5,593	-9,757	12,438
EX-IB	Varianzen sind gleich	,519	,473	,812	4,671	,456	1,340	1,651	-2,995	5,675
	Varianzen sind nicht gleich			,346	99	,730	,245	,707	-1,158	1,648
EX-IG	Varianzen sind gleich	,007	,935	1,989	97,000	,050	,245	,123	,001	,489
	Varianzen sind nicht gleich			-,441	99	,660	-2,660	6,032	-14,628	9,308
	Varianzen sind nicht gleich			-,795	2,457	,496	-2,660	3,348	-14,779	9,460

Tabelle 20-37 Mehrfachvergleich zwischen den Aufgabentypen, Häufigkeiten

Mehrfachvergleiche							
Tamhane							
Abhängige Variable	(I) 1=L, 2=LE, 3=CL	(J) 1=L, 2=LE, 3=CL	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
RE-LI	1	2	5,103	3,716	,443	-4,14	14,35
		3	4,069	4,183	,706	-6,23	14,37
	2	1	-5,103	3,716	,443	-14,35	4,14
		3	-1,034	2,694	,974	-7,71	5,64
	3	1	-4,069	4,183	,706	-14,37	6,23
		2	1,034	2,694	,974	-5,64	7,71
RE-IN	1	2	-20,464	4,439	,000	-31,35	-9,58
		3	-43,596	9,439	,000	-67,38	-19,81
	2	1	20,464	4,439	,000	9,58	31,35
		3	-23,132	9,864	,072	-47,79	1,52
	3	1	43,596	9,439	,000	19,81	67,38
		2	23,132	9,864	,072	-1,52	47,79
RE-TX	1	2	121,779	17,459	,000	-164,83	-78,72
		3	-98,072	20,241	,000	-148,75	-47,40
	2	1	121,779	17,459	,000	78,72	164,83
		3	23,706	24,698	,714	-36,97	84,38
	3	1	98,072	20,241	,000	47,40	148,75
		2	-23,706	24,698	,714	-84,38	36,97
RE-HL	1	2	-,646	1,183	,930	-3,54	2,25
		3	-6,559	3,067	,116	-14,29	1,17
	2	1	-,646	1,183	,930	-2,25	3,54
		3	-5,913	3,099	,183	-13,70	1,88
	3	1	6,559	3,067	,116	-1,17	14,29
		2	5,913	3,099	,183	-1,88	13,70
RE-IX	1	2	2,867	4,613	,900	-8,43	14,17
		3	11,855	3,621	,007	2,81	20,90
	2	1	-2,867	4,613	,900	-14,17	8,43
		3	8,988	3,205	,022	1,05	16,93
	3	1	-11,855	3,621	,007	-20,90	-2,81
		2	-8,988	3,205	,022	-16,93	-1,05
RE-LINK	1	2	-3,620	2,242	,305	-9,21	1,97
		3	,200	,320	,901	-,60	1,00
	2	1	3,620	2,242	,305	-1,97	9,21
		3	3,820	2,224	,256	-1,73	9,37
	3	1	-,200	,320	,901	-1,00	,60
		2	-3,820	2,224	,256	-9,37	1,73
RE-IG	1	2	,788	,788	,692	-1,20	2,77
		3	,305	,862	,979	-1,83	2,44
	2	1	-,788	,788	,692	-2,77	1,20
		3	-,483	,350	,445	-1,37	,40
	3	1	-,305	,862	,979	-2,44	1,83
		2	,483	,350	,445	-,40	1,37
SC-LI	1	2	-1,993	1,004	,149	-4,47	,48
		3	-3,397	1,972	,259	-8,38	1,58
	2	1	1,993	1,004	,149	-,48	4,47
		3	-1,404	2,136	,886	-6,72	3,92
	3	1	3,397	1,972	,259	-1,58	8,38
		2	1,404	2,136	,886	-3,92	6,72
SC-TX	1	2	8,023	5,308	,353	-4,97	21,01
		3	-1,285	7,871	,998	-20,82	18,25
	2	1	-8,023	5,308	,353	-21,01	4,97
		3	-9,309	7,800	,560	-28,67	10,06
	3	1	1,285	7,871	,998	-18,25	20,82
		2	9,309	7,800	,560	-10,06	28,67
SC-IN	1	2	-,928	1,288	,855	-4,10	2,25
		3	-5,858	2,124	,028	-11,21	-,51
	2	1	,928	1,288	,855	-2,25	4,10
		3	-4,930	2,358	,121	-10,77	,91
	3	1	5,858	2,124	,028	,51	11,21
		2	4,930	2,358	,121	-,91	10,77
SC-HL	1	2	-5,054	3,959	,500	-14,75	4,64
		3	-3,803	3,686	,667	-12,88	5,27
	2	1	5,054	3,959	,500	-4,64	14,75
		3	1,251	4,313	,988	-9,32	11,82
	3	1	3,803	3,686	,667	-5,27	12,88
		2	-1,251	4,313	,988	-11,82	9,32
SC-IX	1	2	1,597	3,790	,966	-7,68	10,87

		3	1,086	4,157	,991	-9,14	11,31
	2	1	-1,597	3,790	,966	-10,87	7,68
		3	-,511	4,136	,999	-10,67	9,65
	3	1	-1,086	4,157	,991	-11,31	9,14
		2	-,511	4,136	,999	-9,65	10,67
	1	2	2,939	3,448	,782	-5,60	11,48
		3	3,319	3,294	,685	-4,89	11,53
SC-LINK	2	1	-2,939	3,448	,782	-11,48	5,60
		3	-,379	1,851	,996	-4,16	4,92
	3	1	-3,319	3,294	,685	-11,53	4,89
		2	-,379	1,851	,996	-4,92	4,16
	1	2	-29,219	4,759	,000	-41,09	-17,35
		3	-14,930	3,830	,002	-24,63	-5,23
EX-PI	2	1	29,219	4,759	,000	17,35	41,09
		3	14,289	6,062	,063	-,56	29,14
	3	1	14,930	3,830	,002	5,23	24,63
		2	-14,289	6,062	,063	-29,14	-,56
	1	2	3,233	2,653	,546	-3,42	9,88
		3	1,044	2,871	,978	-6,07	8,16
SC-IB	2	1	-3,233	2,653	,546	-9,88	3,42
		3	-2,189	1,366	,311	-5,59	1,22
	3	1	-1,044	2,871	,978	-8,16	6,07
		2	2,189	1,366	,311	-1,22	5,59
	1	2	-,154	-,154	,691	-,54	-,23
		3	-,621	-,356	,252	-1,52	-,28
EX-IB	2	1	-,154	-,154	,691	-,23	-,54
		3	-,467	-,388	,554	-1,43	-,50
	3	1	-,621	-,356	,252	-,28	1,52
		2	-,467	-,388	,554	-,50	1,43
	1	2	2,683	1,821	,386	-1,89	7,26
		3	-1,796	3,373	,934	-10,14	6,55
EX-IG	2	1	-2,683	1,821	,386	-7,26	1,89
		3	-4,479	2,871	,341	-11,76	2,80
	3	1	1,796	3,373	,934	-6,55	10,14
		2	4,479	2,871	,341	-2,80	11,76

\*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

Tabelle 20-38 T-Test über die Mittelwerte aller drei TT emotionale Lage positiv vs. negativ (summe valence und arousal), absolute Häufigkeiten

Test bei unabhängigen Stichproben											
			Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
			F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
										Untere	Obere
RE-LI	Varianzen gleich	sind	1,773	,188	-,494	64	,623	-2,398	4,851	-12,088	7,293
	Varianzen nicht gleich	sind			-,368	17,911	,717	-2,398	6,513	-16,086	11,291
RE-IN	Varianzen gleich	sind	1,475	,229	1,314	64	,193	10,850	8,255	-5,641	27,341
	Varianzen nicht gleich	sind			1,583	36,744	,122	10,850	6,853	-3,038	24,738
RE-TX	Varianzen gleich	sind	,162	,689	2,712	64	,009	66,315	24,452	17,467	115,163
	Varianzen nicht gleich	sind			2,718	25,433	,012	66,315	24,398	16,110	116,520
RE-HL	Varianzen gleich	sind	3,447	,068	1,652	64	,103	5,490	3,322	-1,147	12,127
	Varianzen nicht gleich	sind			2,408	58,586	,019	5,490	2,280	,927	10,053
RE-IX	Varianzen gleich	sind	,804	,373	-1,057	64	,295	-6,000	5,679	-17,345	5,345
	Varianzen nicht gleich	sind			-,996	23,174	,330	-6,000	6,024	-18,456	6,456
RE-LINK	Varianzen gleich	sind	,552	,460	,402	64	,689	1,257	3,127	-4,990	7,505
	Varianzen nicht gleich	sind			,616	63,078	,540	1,257	2,042	-2,824	5,339
RE-IG	Varianzen gleich	sind	2,755	,102	,814	64	,419	,800	,983	-1,164	2,764



### 8.10. Korrelation der Videolabels

[illegible]

RE-TX	Korrelation nach Pearson	0,168	0,530	0,101	0,031	0,039	0,067	0,023	0,162	0,212	0,057	0,017	0,061	0,159	0,327	0,177	0,152	0,055	0,056	0,176	0,206	0,021	0,157	0,064
	Signifikanz (2-seitig)	0,092	0,000		0,759	0,699	0,507	0,818	0,105	0,033	0,574	0,865	0,547	0,113	0,001	0,077	0,128	0,586	0,575	0,078	0,039	0,837	0,116	0,528
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
RE-HL	Korrelation nach Pearson	0,109	0,150	0,031	0,108	0,115	0,006	0,032	0,313	0,363	0,133	0,159	0,157	0,058	0,039	0,090	0,031	0,190	0,133	0,203	0,182	0,084	0,026	
	Signifikanz (2-seitig)	0,277	0,135	0,759		0,281	0,251	0,955	0,752	0,001	0,000	0,184	0,113	0,118	0,564	0,373	0,758	0,058	0,186	0,042	0,068	0,405	0,794	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
RE-IX	Korrelation nach Pearson	0,156	0,075	0,039	0,108	0,101	0,302	0,034	0,030	0,036	0,150	0,102	0,126	0,195	0,088	0,096	0,102	0,083	0,330	0,043	0,051	0,200	0,305	0,183
	Signifikanz (2-seitig)	0,120	0,458	0,699	0,281		0,002	0,733	0,766	0,723	0,135	0,309	0,209	0,050	0,379	0,341	0,310	0,407	0,001	0,667	0,614	0,045	0,002	0,066
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
RE-LINK	Korrelation nach Pearson	0,079	0,013	0,067	0,115	0,302	0,101	0,021	0,049	0,121	0,051	0,078	0,033	0,078	0,030	0,023	0,037	0,047	0,029	0,078	0,095	0,094	0,135	0,164
	Signifikanz (2-seitig)	0,431	0,896	0,507	0,251	0,002		0,836	0,629	0,229	0,611	0,437	0,746	0,439	0,769	0,817	0,711	0,641	0,773	0,438	0,347	0,351	0,179	0,102
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
RE-IG	Korrelation nach Pearson	0,030	0,008	0,023	0,006	0,034	0,021	0,074	0,026	0,017	0,018	0,041	0,064	0,062	0,012	0,029	0,015	0,015	0,124	0,031	0,039	0,148	0,025	
	Signifikanz (2-seitig)	0,767	0,936	0,818	0,955	0,733	0,836		0,463	0,796	0,866	0,858	0,683	0,526	0,539	0,906	0,777	0,878	0,878	0,218	0,757	0,700	0,140	0,808
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
SC-LI	Korrelation nach Pearson	0,063	0,051	0,162	0,032	0,030	0,049	0,074	0,042	0,062	0,043	0,035	0,009	0,308	0,193	0,077	0,168	0,201	0,011	0,008	0,056	0,104	0,098	
	Signifikanz (2-seitig)	0,534	0,611	0,105	0,752	0,766	0,629	0,463	0,678	0,541	0,669	0,727	0,929	0,002	0,053	0,444	0,094	0,044	0,916	0,940	0,578	0,302	0,331	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
SC-TX	Korrelation nach																							

SC-IB	Korrelation nach Pearson	0,004	0,033	-0,177	-0,039	-0,096	-0,023	-0,012	0,193	0,162	-0,057	-0,129	-0,093	-0,028	-0,010	0,088	0,129	0,017	-0,100	-0,112	0,089	-0,017	-0,078	
	Signifikanz (2-seitig)	0,967	0,740	0,077	0,699	0,341	0,817	0,906	0,053	0,105	0,570	0,200	0,353	0,778	0,919	0,380	0,199	0,868	0,318	0,265	0,378	0,863	0,439	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
EX-IB	Korrelation nach Pearson	-0,028	0,163	0,152	-0,090	-0,102	-0,037	-0,029	0,077	0,102	0,401	0,454	0,283	-0,069	-0,049	0,081	0,206	0,165	-0,071	-0,215	-0,138	-0,060	-0,202	
	Signifikanz (2-seitig)	0,778	0,103	0,128	0,373	0,310	0,711	0,777	0,444	0,312	0,000	0,000	0,004	0,493	0,625	0,380	0,039	0,099	0,483	0,031	0,170	0,552	0,042	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
EX-IG	Korrelation nach Pearson	0,160	0,150	0,055	0,031	0,083	0,047	0,015	0,168	0,068	0,184	0,267	0,101	-0,008	0,039	0,129	0,206	0,041	0,100	-0,044	-0,020	-0,031	-0,119	
	Signifikanz (2-seitig)	0,109	0,133	0,586	0,758	0,407	0,641	0,878	0,094	0,499	0,066	0,007	0,313	0,935	0,699	0,199	0,039	0,683	0,320	0,664	0,840	0,757	0,236	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
Valence	Korrelation nach Pearson	-0,065	-0,125	0,056	-0,190	-0,330	0,029	0,015	-0,201	-0,032	0,055	0,138	-0,057	-0,069	0,073	0,017	0,165	0,041	0,111	-0,143	0,124	0,050	-0,300	-0,213
	Signifikanz (2-seitig)	0,516	0,212	0,575	0,058	0,001	0,773	0,878	0,044	0,753	0,586	0,169	0,569	0,494	0,467	0,868	0,099	0,683	0,155	0,216	0,621	0,002	0,032	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
Arousal	Korrelation nach Pearson	0,025	0,062	0,176	0,133	-0,043	0,078	0,124	-0,011	0,081	0,088	-0,026	-0,003	-0,072	0,118	0,100	0,071	0,100	-0,143	0,111	0,184	-0,120	-0,019	-0,074
	Signifikanz (2-seitig)	0,800	0,537	0,078	0,186	0,667	0,438	0,218	0,916	0,422	0,382	0,796	0,976	0,472	0,239	0,318	0,483	0,320	0,155	0,066	0,231	0,849	0,459	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
Geschlecht (1=w, 2=m)	Korrelation nach Pearson	-0,007	-0,223	-0,206	-0,203	-0,051	0,095	0,031	-0,008	-0,194	-0,058	-0,084	-0,121	-0,171	0,082	0,112	0,215	-0,044	0,124	0,184	0,111	-0,094	-0,097	-0,089
	Signifikanz (2-seitig)	0,948	0,025	0,039	0,042	0,614	0,347	0,757	0,940	0,052	0,567	0,403	0,228	0,087	0,416	0,265	0,031	0,664	0,216	0,066	0,350	0,333	0,377	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
Alter	Korrelation nach Pearson	0,044	0,063	-0,021	-0,182	0,200	0,094	0,039	0,056	0,110	0,010	0,144	-0,009	-0,017	-0,018	0,089	-0,138	-0,020	0,050	-0,120	-0,094	0,111	0,148	0,023
	Signifikanz (2-seitig)	0,660	0,533	0,837	0,068	0,045	0,351	0,700	0,578	0,274	0,918	0,150	0,932	0,864	0,859	0,378	0,170	0,840	0,621	0,231	0,350	0,141	0,817	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
WP-Kenntnis	Korrelation nach Pearson	0,161	0,127	0,157	0,084	0,305	0,135	0,148	0,104	-0,181	0,006	-0,037	0,003	0,022	0,065	0,017	0,060	0,031	-0,300	-0,019	-0,097	0,148	0,111	0,376
	Signifikanz (2-seitig)	0,108	0,207	0,116	0,405	0,002	0,179	0,140	0,302	0,070	0,951	0,710	0,973	0,826	0,519	0,863	0,552	0,757	0,002	0,849	0,333	0,141	0,000	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	
Vorkenntnis_Thema	Korrelation nach Pearson	0,231	-0,135	-0,064	-0,026	0,183	0,164	0,025	-0,098	0,131	0,125	0,158	-0,043	0,082	0,053	0,078	-0,202	-0,119	-0,213	-0,074	-0,089	0,023	0,376	0,111
	Signifikanz (2-seitig)	0,020	0,179	0,528	0,794	0,066	0,102	0,808	0,331	0,192	0,212	0,114	0,671	0,417	0,599	0,439	0,042	0,236	0,032	0,459	0,377	0,817	0,000	
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	